A Riemannian Geometric Approach to Blind Signal Recovery for Grant-Free Radio Network Access

Carlos Feres, *Member, IEEE*, Zhi Ding, *Fellow, IEEE*
Department of Electrical and Computer Engineering
University of California, Davis, CA, USA

***Abstract*—We propose a new nonconvex framework for blind multiple signal demixing and recovery. The proposed Riemannian geometric approach extends the well known constant modulus algorithm to facilitate grant-free wireless access for simultaneous demixing and recovery of multiple signal demixing and recovery. We formulate the problem as non-convex problem optimization problem integrated with the signal orthogonality constraint in the form of Riemannian Orthogonal CMA (ROCMA). Unlike traditional stochastic gradient solutions that require large data samples, parameter tuning, and careful initialization, we leverage Riemannian geometry and transform the orthogonality requirement of recovered signals into a Riemannian manifold optimization. Our solution demonstrates full recovery of multiple access signals without large data sample size or special initialization with high probability of success.**

***Index Terms*—Blind demixing, grant-free access, signal recovery, Riemannian manifolds, non-convex optimization.**

I. Introduction

Recent advances in next generation networking technologies are poised to ubiquitously connect the full spectrum of sensors, devices, and computers to facilitate future development of smart cities and smart agriculture, among other applications. These exciting developments, known collectively as Internet of Things (IoT), promise significant benefits in a plethora of fields including health care, farming, environmental science, infrastructure, energy efficiency, transportation, safety and sustainability. In this work, we focus on the technical challenge of grant-free wireless access for a large number of low complexity wireless devices.

Generally, wireless networks are based on either random access (e.g., WiFi networks) or controlled scheduling (e.g., 4G-LTE cellular networks). Contention based random access schemes, such as the CSMA-CA protocol adopted in IEEE 802.11a/g/n/ac, possess the advantage of simplicity but suffer from lower spectrum efficiency due to access collision when the number of active devices is large. Controlled user scheduling based on centralized access grants can achieve high spectrum efficiency but require more elaborate network-user interaction such as random access and contention resolution, and would incur in higher energy consumption for many low power devices.

A typical IoT application involves sporadic communications between a significant number of transceivers, triggered by external events, in order to save energy. This prompts the need for low-latency communications and the ability to support these links in the performance-constrained scenario of typical IoT transceivers, in particular, in terms of bandwidth efficiency.

In grant-free access, multiple signals could collide at the receiving node. Although these colliding signals can utilize specialized pilots or training signals as their unique characteristics to be exploited for signal separation, there are at least two problems. First, the training signals would consume precious device energy and network bandwidth to transmit but carry no payload data. Second, to reduce training overhead, shorter pilots should be used. However, there are at most N orthogonal training sequences of length N. Thus, a large number of IoT devices pose challenges to both spectrum and energy efficiency.

Blind equalization has been a staple idea in terms of achieving this goal by diminishing the impact of pilots or preambles, aiming to reduce their impact in the overall bandwidth efficiency. Among blind equalization algorithms, the Constant Modulus Algorithm (CMA) presented by Godard [1] in the 1980s is often considered the most widespread technique due to its computational simplicity and practical effectiveness [2], [3]. However, one of its major drawbacks in practical applications is the presence of local extrema -due to the action of additive noise- and its slow convergence [3]–[5]. CMA-based grant-free signal recovery typically apply traditional adaptive algorithms such as stochastic gradient descent to find optimum parameters of an underlying linear system for signal recovery. Such solutions would require finely tuning of e.g., normalization and stepsize, for satisfactory convergence.

There have been extensive works on CMA and other related formulations aimed at overcoming their drawbacks. One interesting approach is the transformation of CMA-based equalization to a convex problem, via semidefinite relaxation [6]–[8], which provides global convergent solutions in a lifted higher dimensional parameter space that are further projected to the original solution space. There are also other relaxation approaches, such as using the trace (nuclear) norm as surrogate for the rank-1 constrain imposed on the CMA problem when defined in terms of matrices [9]. As with any relaxation approach, CMA based on convex relaxation relies on the expectation that the convex problem yields solutions that can be projected to near optimum CMA solutions. Additionally, the problem size grows polynomially with increasing parameter size of the linear system and poses severe practical challenges in many scenarios.

Other line of works include analytical solutions to CMA [10], [11]. These approaches and its variants [12], [13] do have convergence ambiguity owing to the algebraic solution.

This material is based upon work supported by the National Science Foundation under Grants No. 2009001, No. 1711823, and No. 2029027. This work is also supported by the National Agency for Research and Development (ANID) / Scholarship Program / DOCTORADO BECAS CHILE/2016 72170648.

However, they are much more complex and cannot work with QAM source signals that do not exhibit constant modulus (magnitude), such as 16-QAM. There are also multistage schemes [14], that depend heavily on the estimation error being close to the MMSE estimate in earlier stages, or the error accumulates through different stages [15].

Moreover, several CMA-based approaches have been proposed to enforce the recovery of multiple signals at a time. One family of solutions is to consider regularization terms in the cost function [16], [17], which enforce the recovered signals to be uncorrelated. Other schemes propose to modify the iterate after the gradient descent update, such as performing an iterative orthogonality enforcement on the combiners [18]. In general, these approaches are slow, as they require a rather large amount of samples and/or iterations to attain sufficient interference rejection of all recovered sources.

In the current paper, we present a new Riemannian perspective with which we redefine the orthogonality requirement of different demixing combiners as a Riemannian manifold. Any optimization procedure that solves the blind signal recovery problem is formulated as an unconstrained optimization problem over a Riemannian manifold that incorporates the orthogonality of multiple source signals. Leveraging the Riemannian geometry, this is the first known work that directly incorporates the signal orthogonality constraint within the constant modulus optimization problem with established convergence proof. The Riemannian geometric formulation has been extensively studied in recent years [19] and has been successfully applied to several domains, such as low-rank matrix decomposition [20], singular value decomposition [21], phase retrieval [22], blind signal demixing [23], dictionary learning [24], among others. As we shall show, it presents a promising direction for improving CMA and related algorithms for blind signal recovery in grant-free network access.

Section II presents the signal model for blind signal recovery, the formulation of the optimization problem, and some comments on the nature of these. Section III introduces the Riemannian geometry for CMA-based signal recovery, and some optimization schemes. Section IV details theoretical convergence properties and complexity analysis of this technique. Section V presents numerical simulations on each scenario, and finally Section VI summarizes our conclusions.

Notations: In the following, vectors and matrices will be denoted with small and capital boldface letters, such as $\boldsymbol{z}$ and $\boldsymbol{Z}$ respectively. Sets are denoted with calligraphic capital letters. Complex conjugation is denoted with $\overline{z}$. For a complex scalar a, we use $\mathrm{Re}(a)$, $\mathrm{Im}(a)$, $|a|$ and $\angle(a)$ to denote its real part, imaginary part, magnitude and angle, respectively. The transpose, element-wise complex conjugation and conjugate transpose are denoted by $\boldsymbol{z}^{\mathsf{T}}$, $\overline{\boldsymbol{z}}$ and $\boldsymbol{z}^{\mathsf{H}}$, respectively. The Hermitian and skew-Hermitian parts of a matrix $\boldsymbol{Z}$ are denoted as $\mathrm{herm}(\boldsymbol{Z}) = 0.5(\boldsymbol{Z} + \boldsymbol{Z}^{\mathsf{H}})$ and $\mathrm{skew}(\boldsymbol{Z}) = 0.5(\boldsymbol{Z} - \boldsymbol{Z}^{\mathsf{H}})$. The Euclidean norm of vectors and spectral norm of matrices is denoted by $\|\cdot\|$, and the Frobenius norm of matrices is denoted by $\|\cdot\|_F$. Finally, $\mathrm{diag}(\boldsymbol{z})$ represents a diagonal matrix that uses elements of vector $\boldsymbol{z}$ on its diagonal, and $\circ$ denotes the Hadamard (element-wise) product.

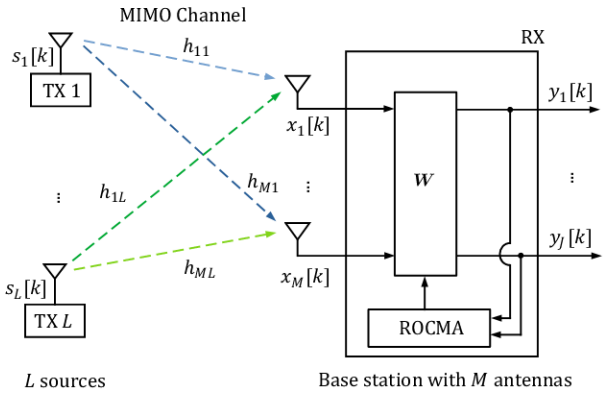


Fig. 1: L sources share a common resource block and transmit independent signals to a host station with M antennas through an unknown physical channel. The host receiver aims to use a linear matrix demixer $\boldsymbol{W}$ to recover $J \leq L$ sources with little mutual interference.

II. System Model

A. Grant-Free Blind Signal Recovery and Demixing

We consider the signal recovery of multiple users in an access group in a grant-free access system, as depicted in Fig. 1. In particular, all potential uplink users in each access group have acquired network timing such that their uplink transmission bursts would span one given set of receiver time slots. Users in each designated access group may randomly transmit within their shared channel in terms of allocated time or frequency resources. Appropriate coding and rate-matching is utilized by all source nodes to have equal number of data symbols K within each access group and burst. Furthermore, we design systems such that with very high probability or certainty that the number of single-antenna active nodes L shall fall below the number of diversity antennas M at the receiver node. In particular, the receiver node does not necessarily know L. Since the receiver recovers multiple user signals during blind demixing without prior knowledge of their identities, the receiver can utilize user-ID scrambled CRC to check which recovered user signal belongs to which user, similar to the blind detection of PDCCH by users using RNTI-scrambled CRC in LTE or 5G [25], [26].

To summarize, we define the received signal vector $\boldsymbol{x}_k$, the transmitted signal vector $\boldsymbol{s_k}$, and the flat fading channel $\boldsymbol{H}$, respectively, as

$$\boldsymbol{x}_k = \begin{bmatrix} x_1[k] \\ \vdots \\ x_M[k] \end{bmatrix}, \; \boldsymbol{s}_k = \begin{bmatrix} s_1[k] \\ \vdots \\ s_L[k] \end{bmatrix}, \; \boldsymbol{H} = \begin{bmatrix} h_{11} & \cdots & h_{1L} \\ \vdots & \cdots & \cdots \\ h_{M1} & \cdots & h_{ML} \end{bmatrix}. \tag{1}$$

Then the received signal vector can be written as

$$\boldsymbol{x}_k = \boldsymbol{H}\boldsymbol{s}_k + \boldsymbol{n}_k, \tag{2}$$

where the MIMO channel matrix $\boldsymbol{H} \in \mathbb{C}^{M \times L}$ is assumed to have full column rank L (with $L \leq M$) and $\boldsymbol{n}_k \in \mathbb{C}^M$ is the vector of additive white Gaussian noises (AWGN) in that resource block, of the same size as $\boldsymbol{x}_k$ in Eq.(1).

In blind multiple signal demixing and recovery, we are interested in deriving J simultaneous demixers $\boldsymbol{w}_\ell \in \mathbb{C}^M$, $\ell \in \{1, \ldots, J\}$ that allow the recovery of J sources with minimal interference, each tuned to a distinct signal. We can also write $\boldsymbol{W} = [\boldsymbol{w}_1 \ \boldsymbol{w}_2 \ \cdots \boldsymbol{w}_\ell]$ as the receiver blind demixing parameter matrix such that

$$\boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{w}_1^{\mathsf{H}} \\ \vdots \\ \boldsymbol{w}_\ell^{\mathsf{H}} \end{bmatrix} \boldsymbol{x}_k = \boldsymbol{W}^{\mathsf{H}} \boldsymbol{x}_k = \begin{bmatrix} \hat{s}_{\ell_1}[k] \\ \vdots \\ \hat{s}_{\ell_J}[k] \end{bmatrix}, \quad \ell_i \in \{1, \ldots, L\}. \tag{3}$$

Note that the receiver has explicit knowledge on neither the unknown channels $\boldsymbol{H}$ nor the number of active sources L, except for the statistical properties and the constellation of each source signal. Additionally, we need to ensure that the demixers do not merely restore the same source signal for only a small subset of sources, possibly with different phases or delays [3]. Therefore, when considering simultaneous multiple signal recovery, additional constraints must be enforced for demixers $\boldsymbol{w}_\ell$, $\ell \in \{1, \ldots, J\}$ to recover different source signals. Without loss of generality, we consider $J \leq L$ and we shall also consider the case when accurate estimation of the number of access users is achieved such that $J = L$.

B. Constant Modulus Algorithm

The problem of blind signal recovery has been extensively studied before. In particular, Godard [1] proposed what was later known [2] as the constant modulus algorithm (CMA) to adaptively find a single optimum demixer $\boldsymbol{w} \in \mathbb{C}^M$ by minimizing the mean CM cost for equalization:

$$E\{[|y_k|^2 - R_2]^2\}, \quad R_2 = \frac{E\{|s_\ell[k]|^4\}}{E\{|s_\ell[k]|^2\}}, \tag{4}$$

where the constant R_2 is computed from the high-order statistics of the source symbols $s(k)$ to match the input-output constellation scale [1]. It is known that CMA can be applied to i.i.d. signals using QAM constellations of arbitrary size and magnitude [3]. Moreover, R_2 can be an arbitrary scalar, e.g. $R_2 = 1$, which will scale $\boldsymbol{w}$ accordingly, and CMA still converge such that its output recovers QAM source signals, without affecting signal integrity. In batch implementation, the single-source CM cost can be rewritten as

$$f(\boldsymbol{w}) = \frac{1}{2K} \sum_{k=1}^{K} \left(|\boldsymbol{x}_k^{\mathsf{H}} \boldsymbol{w}|^2 - R_2 \right)^2, \tag{5}$$

which is a smooth real-valued nonconvex function of $\boldsymbol{w}$. Note that f presents phase invariance, i.e., if $\widehat{\boldsymbol{w}}$ is a solution that minimizes $f(\widehat{\boldsymbol{w}})$, then the entire set $\mathcal{W}(\widehat{\boldsymbol{w}}) = \{\mathrm{e}^{j\theta} \widehat{\boldsymbol{w}} : \theta \in [0, 2\pi]\}$ contains equivalent solutions that achieve the same minimum $f(\widehat{\boldsymbol{w}})$.

C. Simultaneous Multiple Signal Recovery in Demixing

The CMA has been adapted in the past for simultaneous recovery of multiple independent source signals. In these applications, the first step is to define a cumulative demixing cost consisting of J copies of CM costs:

$$f(\boldsymbol{W}) = \frac{1}{2K} \sum_{\ell=1}^{J} \sum_{k=1}^{K} \left(|\boldsymbol{x}_k^{\mathsf{H}} \boldsymbol{w}_\ell|^2 - R_2 \right)^2. \tag{6}$$

The full blind demixing cost is a function of the J blind demixers $\boldsymbol{W}$. The joint blind demixing problem is to optimize multiple solution vectors $\widehat{\boldsymbol{W}} = \begin{bmatrix} \widehat{\boldsymbol{w}}_1 & \cdots & \widehat{\boldsymbol{w}}_J \end{bmatrix}$ that jointly minimize the cumulative CM cost of (6).

This cumulative CM cost by itself cannot guarantee that the recovered signals are indeed from different sources. In fact, even if every one column vector of $\widehat{\boldsymbol{W}}$ captures the same signal source, the cumulative CM cost of (6) is still minimized and cannot prevent such solutions. For this reason, it is clear that the cumulative CM cost of (6) is non-convex and is in fact multi-modal. Hence, the challenge lies in the practical need that distinct source signals be recovered by the J solution vectors of $\widehat{\boldsymbol{W}}$.

Several approaches that aim to enforce the demixers to recover different signals in what is called MIMO blind recovery. Specifically, many works would add regularization term(s) to the cost function (6) to penalize against the recovery of identical signals by more than one solution vectors in $\widehat{\boldsymbol{W}}$. For example, in [16], [17] the authors proposed adding a norm of joint cumulants for such source separation objective. Another MIMO CMA approach [27] uses the real part of equalized signals as regularization.

Despite their demonstrated successes, regularization approaches exhibit some drawbacks. First, the regularizing term typically requires to tune a scalar weight, often by trial and error. There is no performance guarantee under various possible scenarios. Second, different regularization approaches might lead to different solutions and performance, while no solution is consistently better than others. Additionally, regularizing terms often increase the computation complexity as regularized cost functions would either require additional computations or delicate non-convex optimization steps. Finally, regularizing terms proposed in the literature generally are limited to promote pairwise signal orthogonality instead of multi-lateral signal orthogonality, and also require more data samples to successfully suppress mutual interferences.

In our approach based on Riemannian geometry, we enforce signal orthogonality among demixer outputs by directly restricting the solution space. Recall the definition of $\boldsymbol{W}$ as joint demixer matrix. Owing to the phase invariance of the CM cost function, the optimum solution satisfies $\hat{\boldsymbol{W}}^{\mathsf{H}} \boldsymbol{H} = \boldsymbol{P}$ where $\boldsymbol{P} \in \mathbb{C}^{J \times L}$ a generalized permutation matrix whose non-zero entries are all of the form $e^{j\phi}$ instead of being restricted to 1. As a result, at the optimum solution we have $\boldsymbol{P}\boldsymbol{P}^{\mathsf{H}} = \boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix of appropriate size. We therefore write the joint signal recovery constraint as

$$\hat{\boldsymbol{W}}^{\mathsf{H}} \boldsymbol{H} \boldsymbol{H}^{\mathsf{H}} \hat{\boldsymbol{W}} = \boldsymbol{I}. \tag{7}$$

However, the blind receiver node has no knowledge of the channel $\boldsymbol{H}$. Therefore, we can leverage source signal

orthogonality and white noise property to estimate $\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}$ from the sample covariance matrix of the data vectors $\boldsymbol{x}_k$:

$$\boldsymbol{R}_{\boldsymbol{X}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{H}} \xrightarrow{K\to\infty} \mathbb{E}\{\boldsymbol{R}_{\boldsymbol{X}}\} = \boldsymbol{H}\boldsymbol{H}^{\mathsf{H}} + \sigma^2 \boldsymbol{I}. \quad (8)$$

Note that, in the absence of noise, the rank of matrix $\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}} \in \mathbb{C}^{M\times M}$ is L ($L \leq M$), i.e., the rank of $\boldsymbol{H}$. Thus, we formulate the optimization problem for multiple signal recovery as orthogonal constant modulus algorithm (**OCMA**):

$$\min \quad f(\boldsymbol{W}) = \frac{1}{2K}\sum_{k=1}^{K} \left\| (\boldsymbol{W}^{\mathsf{H}}\boldsymbol{X}_k\boldsymbol{W}) \circ \boldsymbol{I} - R_2\boldsymbol{I} \right\|_F^2 \quad (9a)$$

$$\text{s.t.} \quad \boldsymbol{W}^{\mathsf{H}}\boldsymbol{R}_{\boldsymbol{X}}\boldsymbol{W} = \boldsymbol{I}, \quad (9b)$$

where $\boldsymbol{X}_k = \boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{H}}$. Hence, the Euclidean gradient of $f(\boldsymbol{W})$ is

$$\nabla_{\boldsymbol{W}} f(\boldsymbol{W}) = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{X}_k\boldsymbol{W}\big((\boldsymbol{W}^{\mathsf{H}}\boldsymbol{X}_k\boldsymbol{W}) \circ \boldsymbol{I} - R_2\boldsymbol{I}\big), \quad (10a)$$

and for a matrix $\boldsymbol{G}$ of the same size as $\boldsymbol{W}$, the directional derivative of $\nabla_{\boldsymbol{W}} f(\boldsymbol{W})$ in direction $\boldsymbol{G}$ is

$$\begin{aligned} &\mathrm{D}\big(\nabla_{\boldsymbol{W}} f(\boldsymbol{W})\big)[\boldsymbol{G}] \\ &= \frac{1}{K}\sum_{k=1}^{K} \Big(\boldsymbol{X}_k\boldsymbol{W}\big(\boldsymbol{I} \circ (\boldsymbol{W}^{\mathsf{H}}\boldsymbol{X}_k\boldsymbol{G} + \boldsymbol{G}^{\mathsf{H}}\boldsymbol{X}_k\boldsymbol{W})\big) \\ &\qquad + \boldsymbol{X}_k\boldsymbol{G}\big((\boldsymbol{W}^{\mathsf{H}}\boldsymbol{X}_k\boldsymbol{W}) \circ \boldsymbol{I} - R_2\boldsymbol{I}\big)\Big). \end{aligned} \quad (10b)$$

D. Estimating the Number of Active Sources for Demixing

Because of the number of active sources L may vary in practice, the literature has often assumed that L is known. However, in grant-free access, such assumption would not be practical, since, at best, we would only be able to limit the maximum number of simultaneous users according to synchronization and slotted scheduling. Thus we shall first present an approach to estimate the number of active sources.

Given that $\boldsymbol{H}$ has rank $L \leq M$, the sample covariance matrix $\boldsymbol{R}_{\boldsymbol{X}}$ in restriction (9b) is not strictly positive definite in the absence of noise. Thus, the restriction cannot be directly defined as a Riemannian manifold. In noisy scenarios and with several data samples, the sample covariance matrix will likely be positive definite, but could be numerically ill-conditioned with large condition number. However, in both cases we can extract a strictly positive definite matrix from the sample covariance matrix from its rank-L approximation.

We first estimate the number of transmitted signals embedded in the received data via Minka's Laplace method [28], and let the result be L. Let the SVD of the channel matrix $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$, with $\boldsymbol{U} \in \mathcal{U}(M)$, and $\boldsymbol{V} \in \mathcal{U}(L)$, i.e.,

$$\begin{aligned} \boldsymbol{H} &= [\boldsymbol{U}_L\ \boldsymbol{U}_L^{\perp}] \begin{bmatrix} \boldsymbol{\Sigma}_L \\ \boldsymbol{0}_{(M-L)\times L} \end{bmatrix} \boldsymbol{V}^{\mathsf{H}} = \boldsymbol{U}_L\boldsymbol{\Sigma}_L\boldsymbol{V}^{\mathsf{H}}, \\ &\boldsymbol{U}_L \in \mathrm{ST}(M,L), \boldsymbol{U}_L^{\perp} \in \mathrm{ST}(M, M-L), \\ &\boldsymbol{\Sigma}_L = \mathrm{diag}(\sigma_1, \ldots, \sigma_L), \end{aligned} \quad (11)$$

where $\mathrm{ST}(M,L) = \{\boldsymbol{A} \in \mathbb{C}^{M\times L} : \boldsymbol{A}^{\mathsf{H}}\boldsymbol{A} = \boldsymbol{I}_L\}$ is the complex Stiefel manifold of orthonormal L-frames in $\mathbb{C}^M$ [21], [29].

First, consider the noiseless scenario, i.e. $\boldsymbol{R}_{\boldsymbol{X}} = \mathbb{E}\{\boldsymbol{R}_{\boldsymbol{X}}\}$ which equals to

$$\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{U}^{\mathsf{H}} = \boldsymbol{U}_L\boldsymbol{\Sigma}_L\boldsymbol{\Sigma}_L^{\mathsf{H}}\boldsymbol{U}_L^{\mathsf{H}} = \boldsymbol{U}_L\boldsymbol{\Lambda}^{\mathsf{H}}\boldsymbol{U}_L^{\mathsf{H}}. \quad (12)$$

From the above decomposition, we can obtain $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$, but not $\boldsymbol{V}$. Also, note that $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_L\boldsymbol{\Sigma}_L^{\mathsf{H}}$ is diagonal with strictly positive entries because $\boldsymbol{H}$ has full-column rank. Both $\boldsymbol{U}_L, \boldsymbol{U}_L^{\perp}$ are full-column rank.

In the noisy case with infinite samples, we have

$$\begin{aligned} \boldsymbol{H}\boldsymbol{H}^{\mathsf{H}} + \sigma^2\boldsymbol{I} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{U}^{\mathsf{H}} + \sigma^2\boldsymbol{I} \\ &= \boldsymbol{U}\big(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} + \sigma^2\boldsymbol{I}\big)^{\mathsf{H}}\boldsymbol{U}^{\mathsf{H}} = \boldsymbol{U}\boldsymbol{\Lambda}_1\boldsymbol{U}^{\mathsf{H}}, \end{aligned} \quad (13)$$

and we form a diagonal matrix $\boldsymbol{\Lambda}$ using the L largest diagonal components of $\boldsymbol{\Lambda}_1 - \sigma^2\boldsymbol{I}$, corresponding to eigenvector matrix $\boldsymbol{U}_L$. We further define $\boldsymbol{\Sigma}_L = \boldsymbol{\Lambda}^{1/2}$. Note that σ^2 can be estimated by, e.g., averaging the $M-L$ smallest diagonal elements of $\boldsymbol{\Lambda}_1$. This approach is very similar to the so-called probabilistic PCA [30], which obtains the principal components of data and a generative model. However, even though Minka's Laplace method is known for satisfactory performance in the limited sample regime, it relies on the assumption of Gaussian signals and could introduce bias when estimating the number of discrete sources.

Nevertheless, all independent sources contribute with a significant component of the sample covariance matrix, related to its significant eigenvalues, whereas noise will only have minor contributions in other directions as their related eigenvalues are much smaller in high SNR regimes. For an under-estimated L, the L-rank approximation of the sample covariance matrix would likely fail to capture all relevant directions of the channel, leading to mutual signal interference in signal recovery. Therefore, we compute the normalized L-rank approximation error

$$\frac{\left\| \boldsymbol{R}_{\boldsymbol{X}} - \boldsymbol{U}_L\boldsymbol{\Lambda}\boldsymbol{U}_L^{\mathsf{H}} \right\|}{\left\| \boldsymbol{R}_{\boldsymbol{X}} \right\|} \quad (14)$$

for comparison against a preset threshold ϵ_r to decide whether L needs to be increased in a update. We also update the L-rank approximation of the sample covariance matrix. Our test results to be shown later demonstrate the general reliability of this rank estimation method for demixing.

III. ROCMA: A Riemannian Manifold Optimization Framework

The Riemannian framework for optimization on manifolds [19] has gained a lot of attention owing to its capability to handle problems with a real-valued objective function defined on a constrained space,

$$\underset{\boldsymbol{M}\in\mathbb{C}^{m\times n}}{\text{minimize}} \quad f(\boldsymbol{M}) \quad \text{s.t.} \quad \boldsymbol{M} \in \mathcal{M}. \quad (15)$$

Note that the (nonlinear) space $\mathcal{M}$ may not be well-defined in terms of addition, continuity, and/or other properties that are typically exploited by regular optimization approaches in

Euclidean spaces. The main idea is to redefine a constrained optimization in Euclidean space into an unconstrained optimization problem over a manifold. Manifolds are topological spaces that, equipped with a metric, locally resemble Euclidean spaces of equal dimension size, but may be quite different globally. Some manifold examples include spheres, the set of rotations, the set of positive semidefinite matrices, the set of fixed-rank matrices, and Stiefel manifolds, among many others.

In this section, we first obtain a suitable Riemannian manifold representation of the ROCMA problem (9). Next, we further exploit the obtained Riemannian manifold to define a quotient Riemannian manifold, with which we can tackle the phase invariance of the demixers directly in the optimization process.

A. Redefining the Geometry of Signal Recovery

Our goal here is to find a suitable geometry that encodes the orthogonality condition of demixers in the search space of Problem (9). Even with a method to estimate the number of sources L, we derive a general version of the geometry where the receiver attempts to recover $J \leq L$ sources. Considering Eq.(12) in restriction (9b), we have

$$\boldsymbol{I}_J = \hat{\boldsymbol{W}}^{\mathsf{H}} \boldsymbol{H} \boldsymbol{H}^{\mathsf{H}} \hat{\boldsymbol{W}} = \hat{\boldsymbol{W}}^{\mathsf{H}} \boldsymbol{U}_L \boldsymbol{\Sigma}_L \boldsymbol{\Sigma}_L^{\mathsf{H}} \boldsymbol{U}_L^{\mathsf{H}} \hat{\boldsymbol{W}}. \tag{16}$$

By introducing a new variable $\boldsymbol{Y}$ such that

$$\boldsymbol{W} = \boldsymbol{U}_L \boldsymbol{\Sigma}_L^{-1} \boldsymbol{Y}, \tag{17}$$

Eq.(16) yields $\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{Y} = \boldsymbol{I}_J$, which defines the complex Stiefel manifold $\mathrm{ST}(L, J)$. Hence, by means of the transformation (17), we obtain a Riemannian manifold representation of restriction (9b) as $\overline{\mathcal{M}} = \mathrm{ST}(L, J)$ that we can use for optimization purposes. From the solution in terms of $\boldsymbol{Y}$, we obtain the demixer matrix directly by a one-to-one scaling by $\boldsymbol{U}_L \boldsymbol{\Sigma}_L^{-1}$. The variable transformation (17) implies the need to rewrite the cost function, Euclidean gradient, and directional derivatives of the gradient. Defining $\boldsymbol{z}_k = \boldsymbol{\Sigma}_L^{-1} \boldsymbol{U}_L^{\mathsf{H}} \boldsymbol{x}_k$ and $\boldsymbol{Z}_k = \boldsymbol{z}_k \boldsymbol{z}_k^{\mathsf{H}} = \boldsymbol{\Sigma}_L^{-1} \boldsymbol{U}_L^{\mathsf{H}} \boldsymbol{X}_k \boldsymbol{U}_L \boldsymbol{\Sigma}_L^{-1}$, we have a new cost function

$$\overline{\overline{g}}(\boldsymbol{Y}) = \frac{1}{2K} \sum_{k=1}^{K} \left\| (\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{Z}_k \boldsymbol{Y}) \circ \boldsymbol{I} - R_2 \boldsymbol{I} \right\|_F^2, \tag{18}$$

whose Euclidean gradient is

$$\nabla_{\boldsymbol{Y}} \overline{\overline{g}}(\boldsymbol{Y}) = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{Z}_k \boldsymbol{Y} \big((\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{Z}_k \boldsymbol{Y}) \circ \boldsymbol{I} - R_2 \boldsymbol{I} \big), \tag{19}$$

and the directional derivative of (19) in direction $\boldsymbol{G}$ is

$$\begin{aligned} &\mathrm{D}\big(\nabla_{\boldsymbol{Y}} \overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{G}] = \\ &\quad \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{Z}_k \boldsymbol{Y} \big((\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{Z}_k \boldsymbol{G} + \boldsymbol{G}^{\mathsf{H}} \boldsymbol{Z}_k \boldsymbol{Y}) \circ \boldsymbol{I} \big) \\ &\qquad + \boldsymbol{Z}_k \boldsymbol{G} \big((\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{Z}_k \boldsymbol{Y}) \circ \boldsymbol{I} - R_2 \boldsymbol{I} \big). \end{aligned} \tag{20}$$

To optimize $\boldsymbol{Y}$ over $\overline{\mathcal{M}}$, we need to first define the linear space that approximates the manifold around a point $\boldsymbol{Y}$, which is called the tangent space at $\boldsymbol{Y}$ and is denoted as $\mathrm{T}_{\boldsymbol{Y}} \overline{\mathcal{M}}$. For $\overline{\mathcal{M}} = \mathrm{ST}(L, J)$, the tangent space is

$$\begin{aligned} \mathrm{T}_{\boldsymbol{Y}} \overline{\mathcal{M}} = \{ &\boldsymbol{E} \in \mathbb{C}^{L \times J} : \boldsymbol{E} = \boldsymbol{Y} \boldsymbol{\Omega} + \boldsymbol{Y}_\perp \boldsymbol{A}, \\ &\boldsymbol{\Omega} = -\boldsymbol{\Omega}^{\mathsf{H}} \in \mathbb{C}^{J \times J}, \boldsymbol{A} \in \mathbb{C}^{(L-J) \times J} \}. \end{aligned} \tag{21}$$

In other words, $\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{E} = \boldsymbol{\Omega}$ is skew-Hermitian $\forall \boldsymbol{E} \in \mathrm{T}_{\boldsymbol{Y}} \overline{\mathcal{M}}$. Its orthogonal complement is known as *normal* space $\mathrm{N}_{\boldsymbol{Y}} \overline{\mathcal{M}}$ and it is given by

$$\mathrm{N}_{\boldsymbol{Y}} \overline{\mathcal{M}} = \{ \boldsymbol{Y} \boldsymbol{A},\ \boldsymbol{A} = \boldsymbol{A}^{\mathsf{H}} \in \mathbb{C}^{J \times J} \}. \tag{22}$$

We can now define length in the tangent space with a Riemannian metric $d_{\boldsymbol{Y}}$, which is a smooth inner product defined at each element $\boldsymbol{Y}$ for elements of the tangent space $\mathrm{T}_{\boldsymbol{Y}} \overline{\mathcal{M}}$. Here, we use the real-trace inner product, given by

$$d_{\boldsymbol{Y}}(\boldsymbol{E}, \boldsymbol{C}) = \mathrm{Re}\big(\mathrm{Tr}(\boldsymbol{E}^{\mathsf{H}} \boldsymbol{C})\big), \quad \boldsymbol{E}, \boldsymbol{C} \in \mathrm{T}_{\boldsymbol{Y}} \overline{\mathcal{M}}. \tag{23}$$

We also define a projection to the tangent space, which allows to restrict optimization only in the directions of interest, which indeed belong to the tangent space. For $\boldsymbol{G} \in \mathbb{C}^{L \times J}$, the projection operator is

$$\mathrm{Proj}_{\boldsymbol{Y}}^{\mathrm{T}}(\boldsymbol{G}) = \boldsymbol{G} - \boldsymbol{Y} \mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}} \boldsymbol{G}\big) \in \mathrm{T}_{\boldsymbol{Y}} \overline{\mathcal{M}}, \tag{24}$$

which enables us to define the Riemannian gradient and Riemannian Hessian from the Euclidean gradient and its directional derivative, respectively.

For optimization purposes, the motion along the manifold from point $\boldsymbol{Y}$ in a given direction $\boldsymbol{E}$ is given by a retraction $\mathrm{R}_{\boldsymbol{Y}}^{\mathrm{St}}(\boldsymbol{E})$, which in our case we select as the polar retraction for the complex Stiefel manifold [29]

$$\begin{aligned} \mathrm{R}_{\boldsymbol{Y}}^{\mathrm{St}}(\boldsymbol{E}) &= (\boldsymbol{Y} + \boldsymbol{E}) \big((\boldsymbol{Y} + \boldsymbol{E})^{\mathsf{H}} (\boldsymbol{Y} + \boldsymbol{E}) \big)^{-0.5} \\ &= (\boldsymbol{Y} + \boldsymbol{E}) \big(\boldsymbol{I} + \boldsymbol{E}^{\mathsf{H}} \boldsymbol{E} \big)^{-0.5}. \end{aligned} \tag{25}$$

B. Riemannian Quotient Geometry

In the context of Riemannian manifold optimization, quotient Riemannian manifolds are used to define a manifold that presents invariance of the cost function or the representation of the manifold itself [31]. It can be defined by equipping the original or ambient manifold with an equivalence relation between its points to describe the aforementioned invariance.

Let $\sim$ be such an equivalence relation, i.e., $\boldsymbol{Y} \sim \boldsymbol{Y}_0$ denotes that $\boldsymbol{Y}$ and $\boldsymbol{Y}_0$ are equivalent in terms of the invariance of interest. Thus, we can identify equivalent points to $\boldsymbol{Y}$ as one single set known as equivalence class, denoted as

$$[\boldsymbol{Y}] = \{ \boldsymbol{Y}_0 \in \mathcal{M} : \boldsymbol{Y}_0 \sim \boldsymbol{Y} \}. \tag{26}$$

When the quotient space of the ambient manifold $\overline{\mathcal{M}}$ under the equivalence relation $\sim$ satisfies certain conditions [19, Chapter 3], it is a Riemannian quotient manifold, which is the set of equivalence classes:

$$\mathcal{M} = \overline{\mathcal{M}} / \sim = \big\{ [\boldsymbol{Y}] : \boldsymbol{Y} \in \overline{\mathcal{M}} \big\}. \tag{27}$$

A quotient manifold is an abstract space whose elements are subsets of the ambient manifold. However, the use of quotient manifolds in Riemannian optimization has additional advantages, such as the ability of obtaining a strictly positive definite

Hessian by neglecting directions related to the cost function invariance, and potential reduction of problem dimensionality by applying a simple representation of the elements in the space. Even in a case when there is no such representation and the ambient manifold is used for computational purposes, the quotient geometry is theoretically important to establish convergence properties of second-order methods that rely on the positive definiteness of the Hessian on the manifold.

Now, recall that the cost function (9a) presents permutation invariance in the demixers, and unimodular phase invariance in each demixer such that for each demixer $\boldsymbol{w}_\ell$, a rotated demixer $\mathrm{e}^{j\theta}\boldsymbol{w}_\ell, \theta \in [0, 2\pi]$ yields the same cost value. However, the permutation group is discrete, and thus is a Lie group of dimension 0. This means that in terms of local behavior, the equivalence classes of permutations behaves exactly the same as the ambient manifold, and thus we dismiss the permutations in our analysis and focus exclusively in the phase invariance.

When considering multiple demixers in $\boldsymbol{W}$, we want to describe unimodular phase invariance on each of the J demixers simultaneously. Let $\mathcal{U}(1)^{\times J}$ the group of diagonal unitary matrices of size J, i.e.

$$\mathcal{U}(1)^{\times J} = \Big\{ \boldsymbol{D} \in \mathcal{U}(J) : \boldsymbol{D} = \operatorname{diag}\left(\mathrm{e}^{j\theta_1} \ \cdots \ \mathrm{e}^{j\theta_J}\right) \Big\}.$$

Thus, the group action of $\mathcal{U}(1)^{\times J}$ defines an equivalence relation between demixer matrices. The corresponding equivalence class is then

$$[\boldsymbol{W}] = \big\{ \boldsymbol{W}\boldsymbol{D} : \boldsymbol{D} \in \mathcal{U}(1)^{\times J} \big\},$$

and by means of (17), we have that $\boldsymbol{W}\boldsymbol{D} = \boldsymbol{U}_L \boldsymbol{\Sigma}_L^{-1} \boldsymbol{Y}\boldsymbol{D}$, i.e. the equivalence class in terms of $\boldsymbol{Y}$ is

$$[\boldsymbol{Y}] = \big\{ \boldsymbol{Y}\boldsymbol{D} : \boldsymbol{D} \in \mathcal{U}(1)^{\times J} \big\}, \tag{28}$$

and we obtain a quotient space that considers the cost function invariance as

$$\mathcal{M} = \overline{\mathcal{M}} / \mathcal{U}(1)^{\times J} \tag{29}$$

with a canonical projection $\pi : \overline{\mathcal{M}} \to \mathcal{M} : \boldsymbol{Y} \mapsto [\boldsymbol{Y}]$.

Furthermore, $\mathcal{M}$ is indeed a Riemannian quotient manifold (which ensures π is smooth). This follows from the fact that the group action of diagonal unitary matrices is smooth, free and proper [31, Section 9.2]: (a) it is smooth because it is a matrix multiplication; (b) it is free because $\boldsymbol{Y}\boldsymbol{D} = \boldsymbol{Y}$ implies $\boldsymbol{D} = \boldsymbol{I}$ by left multiplication of $\boldsymbol{Y}^{\mathsf{H}}$; and (c) it is proper because $\mathcal{U}(1)^{\times J}$ is the maximal torus of the unitary group (a compact group), and thus is compact. Hence, the equivalence classes all have dimension J and are closed embedded submanifolds of $\mathcal{M}$. This quotient manifold has been introduced before in the context of subspace estimation [32], but to the best of our knowledge it has not been developed or used in Riemannian optimization.

The quotient manifold $\mathcal{M}$ is an abstract space, and requires matrix representations in the computational space $\overline{\mathcal{M}}$. Fortunately, an (abstract) element $\boldsymbol{Y}_q$ on the quotient manifold can be represented by an element $\boldsymbol{Y}$ in the computational space. Thus, every geometry-related operation over the quotient manifold can be defined in terms of elements and operations in the computational space.

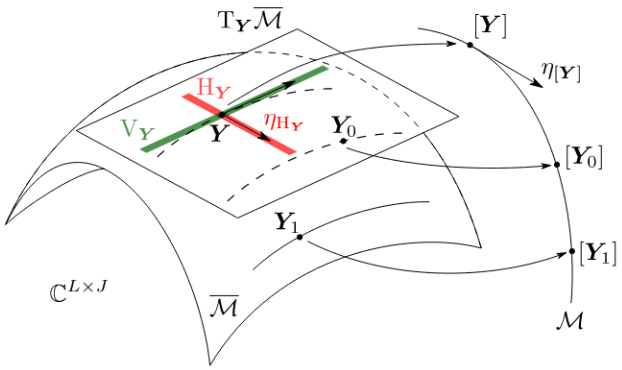


Fig. 2: Representation of the ambient manifold $\overline{\mathcal{M}}$ and quotient manifold $\mathcal{M}$. The tangent space $\mathrm{T}_{\boldsymbol{Y}}\overline{\mathcal{M}}$ is divided into a vertical space $\mathrm{V}_{\boldsymbol{Y}}$ (in green) and a horizontal space $\mathrm{H}_{\boldsymbol{Y}}$ (in red), which contains the relevant search directions $\eta_{\mathrm{H}_{\boldsymbol{Y}}}$. These directions correspond to tangent directions $\eta_{[\boldsymbol{Y}]}$ at the point $[\boldsymbol{Y}]$ in the quotient manifold.

In particular, we look for a representation of the tangent space of the quotient manifold $\mathrm{T}_{\boldsymbol{Y}}\mathcal{M}$ using tangent vectors of the ambient manifold $\overline{\mathcal{M}}$. We accomplish this by characterizing the tangent space $\mathrm{T}_{\boldsymbol{Y}}\overline{\mathcal{M}}$ as the direct sum of two orthogonal spaces: the *vertical* space $\mathrm{V}_{\boldsymbol{Y}}$, which contains the directions tangent to the equivalence classes, and the *horizontal* space $\mathrm{H}_{\boldsymbol{Y}}$, which contains the tangent directions orthogonal to the vertical space. That is, the horizontal space contains the directions of interest in terms of optimization. Hence, an (abstract) tangent vector of the quotient manifold $\boldsymbol{\eta}_{[\boldsymbol{Y}]}$ can be represented as an horizontal vector $\boldsymbol{\eta}_{\mathrm{H}_{\boldsymbol{Y}}}$ of the ambient manifold, which is called the *horizontal lift*, defined as $\boldsymbol{\eta}_{\mathrm{H}_{\boldsymbol{Y}}} = \operatorname{lift}_{\boldsymbol{Y}}(\boldsymbol{\eta}_{[\boldsymbol{Y}]})$. When $\mathcal{M}$ is endowed with the Riemannian metric inherited from $\overline{\mathcal{M}}$,

$$d_{[\boldsymbol{Y}]}(\boldsymbol{\eta}, \boldsymbol{\xi}) = d_{\boldsymbol{Y}}\big(\operatorname{lift}_{\boldsymbol{Y}}(\boldsymbol{\eta}), \operatorname{lift}_{\boldsymbol{Y}}(\boldsymbol{\xi})\big), \ \boldsymbol{\eta}, \boldsymbol{\xi} \in \mathrm{T}_{[\boldsymbol{Y}]}\mathcal{M},$$

the canonical projection π forms a Riemannian submersion from the quotient manifold to the computational space, thereby defining a correspondence between elements of $\mathcal{M}$ and elements of $\overline{\mathcal{M}}$ [19]. Fig. 2 shows a depiction of the quotient manifold geometry and its relation to the ambient manifold.

We now define the vertical and horizontal spaces. Let $\boldsymbol{D} : \mathbb{R} \to \mathcal{U}(1)^{\times J}$ be a path in the equivalence class such that $\boldsymbol{D}(0) = \boldsymbol{I}$. The vertical space is given by vectors of the form $\boldsymbol{Y}\boldsymbol{D}'(0)$ where vectors $\boldsymbol{D}'(t)$ are tangent to $\mathcal{U}(1)^{\times J}$, whose tangent set corresponds to the Lie algebra of unitary diagonal matrices $\mathfrak{t}(J)$, consisting of diagonal imaginary matrices of size $J \times J$. Therefore,

$$\begin{aligned} \mathrm{V}_{\boldsymbol{Y}} &= \big\{ \boldsymbol{Y}\boldsymbol{T} : \boldsymbol{T} \in \mathfrak{t}(J) \big\} \\ &= \{ \boldsymbol{Y}\boldsymbol{T} : \boldsymbol{T} \in \mathbb{C}^{J\times J} \text{ imaginary diagonal} \}, \end{aligned} \tag{30}$$

and the horizontal space is then given by

$$\begin{aligned} \mathrm{H}_{\boldsymbol{Y}} &= \big(\mathrm{V}_{\boldsymbol{Y}}\big)^{\perp} \\ &= \{ \boldsymbol{E} \in \mathrm{T}_{\boldsymbol{Y}}\mathcal{M} : \langle \boldsymbol{E}, \boldsymbol{F} \rangle = 0 \quad \forall \boldsymbol{F} \in \mathrm{V}_{\boldsymbol{Y}} \} \\ &= \{ \boldsymbol{E} \in \mathrm{T}_{\boldsymbol{Y}}\mathcal{M} : \langle \boldsymbol{E}, \boldsymbol{Y}\boldsymbol{T} \rangle = 0 \quad \forall \boldsymbol{T} \in \mathfrak{t}(J) \} \\ &= \{ \boldsymbol{E} \in \mathrm{T}_{\boldsymbol{Y}}\mathcal{M} : \operatorname{Re}\big(\operatorname{Tr}(\boldsymbol{E}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{T})\big) = 0 \quad \forall \boldsymbol{T} \in \mathfrak{t}(J) \} \end{aligned} \tag{31}$$

and thus $\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{E}$ is skew-Hermitian with zero diagonal, to be orthogonal to any $\boldsymbol{T} \in \mathfrak{t}(J)$. This is equivalent to state that the projection to horizontal space is given by

$$\begin{aligned}\mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}(\boldsymbol{G}) &= \mathrm{Proj}^{\mathrm{T}}_{\boldsymbol{Y}}(\boldsymbol{G}) - \boldsymbol{Y}\Big(\boldsymbol{I} \circ \big(\boldsymbol{Y}^{\mathsf{H}}\mathrm{Proj}^{\mathrm{T}}_{\boldsymbol{Y}}(\boldsymbol{G})\big)\Big) \\ &= \boldsymbol{G} - \boldsymbol{Y}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{G}\big) - \boldsymbol{Y}\big(\boldsymbol{I} \circ \mathrm{skew}(\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{G})\big). \quad (32)\end{aligned}$$

Finally, we can inherit the retraction from the ambient manifold via lifting and enforcing the equivalence class. Indeed, for $\boldsymbol{E} = \mathrm{lift}_{\boldsymbol{Y}}(\boldsymbol{\eta})$, we can see that the polar retraction depends only on the equivalence class:

$$\begin{aligned}\mathrm{R}_{[\boldsymbol{Y}]}(\boldsymbol{\eta}) &= \mathrm{R}^{\mathrm{St}}_{\boldsymbol{Y}\boldsymbol{D}}(\boldsymbol{E}) \\ &= (\boldsymbol{Y}\boldsymbol{D} + \boldsymbol{E}\boldsymbol{D})\big(\boldsymbol{I} + (\boldsymbol{E}\boldsymbol{D})^{\mathsf{H}}\boldsymbol{E}\boldsymbol{D}\big)^{-0.5} \\ &= \mathrm{R}^{\mathrm{St}}_{\boldsymbol{Y}}(\boldsymbol{E})\boldsymbol{D} = \big[\mathrm{R}^{\mathrm{St}}_{\boldsymbol{Y}}(\boldsymbol{E})\big]. \quad (33)\end{aligned}$$

Consequently, we can effectively optimize over the quotient manifold $\mathcal{M}$ using representatives from the ambient manifold $\overline{\mathcal{M}}$. Moreover, as $\overline{\mathcal{M}}$ is an embedded submanifold of Euclidean space, we can obtain the Riemannian gradient and Hessian in terms of objects in Euclidean space, which is convenient as we already know the cost function $\overline{\overline{g}}$ in Euclidean space given by (18) with corresponding derivatives (19) and (20). We denote $\overline{g} = \overline{\overline{g}}\big|_{\overline{\mathcal{M}}}$ the restriction of the Euclidean cost function to the ambient manifold. As $\overline{g}$ is invariant under $\sim$, the cost function g defined in $\mathcal{M}$ is smooth and $g([\boldsymbol{Y}]) = \overline{g}(\boldsymbol{Y})$. Hence, by definition, the Riemannian gradient of g is zero for vertical vectors (as they are invariance directions), and we have

$$\begin{aligned}\mathrm{lift}_{\boldsymbol{Y}}\big(\mathrm{grad}_{[\boldsymbol{Y}]}g([\boldsymbol{Y}])\big) &= \mathrm{grad}_{\boldsymbol{Y}}\overline{g}(\boldsymbol{Y}) \\ &= \mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big) = \mathrm{Proj}^{\mathrm{T}}_{\boldsymbol{Y}}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big), \quad (34)\end{aligned}$$

where we can choose the most convenient projection to obtain the lifted gradient, thus we use the projection to tangent space:

$$\begin{aligned}\mathrm{lift}_{\boldsymbol{Y}}\big(\mathrm{grad}_{[\boldsymbol{Y}]}g([\boldsymbol{Y}]\big) &= \mathrm{Proj}^{\mathrm{T}}_{\boldsymbol{Y}}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big) \\ &= \nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y}) - \boldsymbol{Y}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big) \\ &= \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{Z}_k\boldsymbol{Y}\big((\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Z}_k\boldsymbol{Y}) \circ \boldsymbol{I} - R_2\boldsymbol{I}\big) \quad (35) \\ &\quad - \boldsymbol{Y}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Z}_k\boldsymbol{Y}\big((\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Z}_k\boldsymbol{Y}) \circ \boldsymbol{I} - R_2\boldsymbol{I}\big)\big).\end{aligned}$$

The horizontal lift of the Riemannian Hessian of g in terms of $\overline{\overline{g}}$ is given by

$$\begin{aligned}\mathrm{lift}_{\boldsymbol{Y}}\big(\mathrm{Hess}_{[\boldsymbol{Y}]}g([\boldsymbol{Y}])[\boldsymbol{\eta}]\big) &= \mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\big(\mathrm{Hess}_{\boldsymbol{Y}}\overline{g}(\boldsymbol{Y})[\boldsymbol{E}]\big) \\ &= \mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\big(\mathrm{D}\overline{\overline{r}}(\boldsymbol{Y})[\boldsymbol{E}]\big), \quad (36)\end{aligned}$$

with $\boldsymbol{E} = \mathrm{lift}_{\boldsymbol{Y}}(\boldsymbol{\eta})$ and $\overline{\overline{r}}$ a smooth extension of $\mathrm{grad}_{\boldsymbol{Y}}\overline{g}$ to a neighborhood of $\overline{\mathcal{M}}$ in Euclidean space $\mathbb{C}^{L\times J}$. An obvious choice is

$$\begin{aligned}\overline{\overline{r}}(\boldsymbol{Y}) &= \nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y}) - \boldsymbol{Y}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big) \\ &= \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{Z}_k\boldsymbol{Y}\big((\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Z}_k\boldsymbol{Y}) \circ \boldsymbol{I} - R_2\boldsymbol{I}\big) \\ &\quad - \boldsymbol{Y}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Z}_k\boldsymbol{Y}\big((\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Z}_k\boldsymbol{Y}) \circ \boldsymbol{I} - R_2\boldsymbol{I}\big)\big),\end{aligned}$$

with directional derivatives

$$\begin{aligned}\mathrm{D}\overline{\overline{r}}(\boldsymbol{Y})[\boldsymbol{G}] &= \mathrm{D}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{G}] - \boldsymbol{G}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big) \\ &\quad - \boldsymbol{Y}\mathrm{herm}\big(\mathrm{D}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{G}]\big),\end{aligned}$$

and the lifted Riemannian Hessian is

$$\begin{aligned}&\mathrm{lift}_{\boldsymbol{Y}}\big(\mathrm{Hess}_{[\boldsymbol{Y}]}g([\boldsymbol{Y}])[\boldsymbol{\eta}]\big)\mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\Big(\mathrm{D}\overline{\overline{r}}(\boldsymbol{Y})[\boldsymbol{E}]\Big) \\ &\quad= \mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\Big(\mathrm{D}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{E}] - \boldsymbol{E}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big) \\ &\qquad - \boldsymbol{Y}\mathrm{herm}\big(\mathrm{D}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{E}]\big)\Big) \\ &\quad= \mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\Big(\mathrm{D}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{E}] - \boldsymbol{E}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)\Big), \quad (37)\end{aligned}$$

where the last term of the first equality vanishes through the horizontal projection, because the term belongs to the normal space and $\mathrm{H}_{\boldsymbol{Y}} \subset \mathrm{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} \perp \mathrm{N}_{\boldsymbol{Y}}\overline{\mathcal{M}}$.

Table I summarizes the geometric definitions of the quotient manifold $\mathcal{M}$ (using representatives in $\overline{\mathcal{M}}$) for ROCMA. Readers interested in additional details of the quotient manifold discussions may refer to [31, Chapter 9].

C. Riemannian Optimization for Blind Signal Recovery

We use a Riemannian Trust-Region (RTR) algorithm, which is a second-order optimization approach with superlinear convergence rate [33]. At each iteration, to search a direction $\boldsymbol{E}$ in the horizontal space $\mathrm{H}_{\boldsymbol{Y}}$ of iterate $\boldsymbol{Y} \in \overline{\mathcal{M}}$, RTR solves the trust-region subproblem

$$\mathcal{T}: \begin{array}{ll}\underset{\boldsymbol{E}\in\mathrm{H}_{\boldsymbol{Y}}}{\text{minimize}} & q_{\boldsymbol{Y}}(\boldsymbol{E}) \\ \text{s.t.} & d_{\boldsymbol{Y}}(\boldsymbol{E},\boldsymbol{E}) \le \varepsilon^2\end{array} \quad (38)$$

in the ambient manifold, where $q_{\boldsymbol{Y}}$ is a quadratic model of the cost function at $\boldsymbol{Y} \in \overline{\mathcal{M}}$, and ε denotes the trust region radius. The model is given by

$$\begin{aligned}q_{\boldsymbol{Y}}(\boldsymbol{E}) &= \overline{g}(\boldsymbol{Y}) + d_{\boldsymbol{Y}}\Big(\boldsymbol{E}, \mathrm{grad}_{\boldsymbol{Y}}\overline{g}(\boldsymbol{Y})\big)\Big) \\ &\quad + \frac{1}{2}d_{\boldsymbol{Y}}\Big(\boldsymbol{E}, \mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\big(\mathrm{Hess}_{\boldsymbol{Y}}\overline{g}(\boldsymbol{Y})[\boldsymbol{E}]\big)\Big),\end{aligned}$$

using the horizontal lifts of both Riemannian gradient and Hessian derived in Eqs.(34)-(37). In its general formulation, RTR uses a self-adjoint linear operator instead of the Riemannian Hessian for ease of computation, but can obtain a better performance match using the Hessian.

We can now define the ROCMA algorithm as summarized in Algorithm 1. Succinctly, we first initialize by estimating the number of sources L, perform an L-rank eigendecomposition that removes noise contribution in eigenvalues, and by corroborating that the L-rank approximation is close to the sample covariance matrix to adjust L if needed. After scaling the data vectors, we define cost function, quotient Riemannian manifold, and geometry operations. Thereafter, we determine Riemannian Trust Regions: in each iteration we solve the trust-regions subproblem $\mathcal{T}$ in the horizontal space of the current iterate, obtaining a descent direction $\boldsymbol{E}$ in the horizontal space $\mathrm{H}_{\boldsymbol{Y}}$, whose magnitude is given by the size of the accepted trust region [19]. The subsequent solution iterate is computed

TABLE I: Riemannian geometry definitions required for manifold optimization of ROCMA.

Name	Definition
Computational space $\overline{\mathcal{M}}$	$\mathrm{ST}(L,J)$
Quotient space $\mathcal{M}=\overline{\mathcal{M}}/\sim$	$\mathrm{ST}(L,J)/\mathcal{U}(1)^{\times J}$
Horizontal space $\mathrm{H}_{\boldsymbol{Y}}$	$\mathrm{H}_{\boldsymbol{Y}}=\{\boldsymbol{E}\in \mathrm{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} : \mathrm{Re}\big(\mathrm{Tr}(\boldsymbol{E}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{T})\big)=0 \quad \forall \boldsymbol{T}\in\mathfrak{t}(J)\}$
Horizontal space projection $\mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}$	$\mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}(\boldsymbol{G})=\boldsymbol{G}-\boldsymbol{Y}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{G}\big)-\boldsymbol{Y}\big(\boldsymbol{I}\circ\mathrm{skew}(\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{G})\big)$
Riemannian metric $d_{\boldsymbol{Y}}$	$d_{\boldsymbol{Y}}(\boldsymbol{E},\boldsymbol{C})=\mathrm{Re}\big(\mathrm{Tr}(\boldsymbol{E}^{\mathsf{H}}\boldsymbol{C})\big),\ \boldsymbol{E},\boldsymbol{C}\in\mathrm{H}_{\boldsymbol{Y}}$
Retraction $\mathrm{R}^{\mathrm{St}}_{\boldsymbol{Y}}$	$\mathrm{R}^{\mathrm{St}}_{\boldsymbol{Y}}(\boldsymbol{E})=(\boldsymbol{Y}+\boldsymbol{E})\big(\boldsymbol{I}+\boldsymbol{E}^{\mathsf{H}}\boldsymbol{E}\big)^{-0.5}$
Riemannian gradient $\mathrm{grad}_{[\boldsymbol{Y}]}g$	$\mathrm{lift}_{\boldsymbol{Y}}\big(\mathrm{grad}_{[\boldsymbol{Y}]}g([\boldsymbol{Y}]\big)=\mathrm{Proj}^{\mathrm{T}}_{\boldsymbol{Y}}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)$
Riemannian Hessian $\mathrm{Hess}_{[\boldsymbol{Y}]}g$	$\mathrm{lift}_{\boldsymbol{Y}}\big(\mathrm{Hess}_{[\boldsymbol{Y}]}g(\boldsymbol{Y})[\boldsymbol{E}]\big)=\mathrm{Proj}^{\mathrm{H}}_{\boldsymbol{Y}}\Big(\mathrm{D}\big(\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)[\boldsymbol{E}]-\boldsymbol{E}\mathrm{herm}\big(\boldsymbol{Y}^{\mathsf{H}}\nabla_{\boldsymbol{Y}}\overline{\overline{g}}(\boldsymbol{Y})\big)\Big)$

using the retraction of $\boldsymbol{E}$, which brings the result back to the manifold. Once the algorithm converges, we compute the demixer matrix by scaling the obtained solution with $\boldsymbol{U}_L\boldsymbol{\Sigma}_L^{-1}$.

Algorithm 1 Riemannian Orthogonal CMA (ROCMA)

Given: $\boldsymbol{x}_k\in\mathbb{C}^M$, $k\in\{1,\dots,K\}$, trust region radius ε, low-rank approximation tolerance ϵ_r

A) Source estimation:

Estimate number of independent sources L with Minka's Laplace method

Obtain L largest eigenvalues and corresponding eigenvectors of sample covariance matrix $\boldsymbol{R}_{\boldsymbol{X}}$ to construct L-rank approximation $\boldsymbol{R}_{\boldsymbol{X}}=\sum_k \boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{H}}\approx\boldsymbol{U}_L\boldsymbol{\Lambda}\boldsymbol{U}_L^{\mathsf{H}}$

while $\|\boldsymbol{R}_{\boldsymbol{X}}-\boldsymbol{U}_L\boldsymbol{\Lambda}\boldsymbol{U}_L^{\mathsf{H}}\|>\epsilon_r\|\boldsymbol{R}_{\boldsymbol{X}}\|$ **do**

  $L=L+1$

  Update $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}_L$ and $\boldsymbol{U}_L$ with new components

end while

B) Initialization:

Define $\boldsymbol{z}_k=\boldsymbol{\Sigma}_L^{-1}\boldsymbol{U}_L^{\mathsf{H}}\boldsymbol{x}_k$ and objective function g

Define geometry ingredients of $\mathcal{M}$ with representatives in $\overline{\mathcal{M}}$ according to Table I

C) Riemannian Trust Regions:

while not converged **do**

  Obtain descent direction $\boldsymbol{E}_t$ by solving $\mathcal{T}$ in $\mathrm{H}_{\boldsymbol{Y}_t}$

  $\boldsymbol{Y}_{t+1}=\mathrm{R}^{\mathrm{St}}_{\boldsymbol{Y}_t}\big(\boldsymbol{E}_t\big)$

end while

$\boldsymbol{W}_{\mathrm{final}}=\boldsymbol{U}_L\boldsymbol{\Sigma}_L^{-1}\boldsymbol{Y}_{\mathrm{final}}$

A known algorithm to solve the trust-region subproblem $\mathcal{T}$ based in a truncated Conjugate Gradient approach is available as Algorithm 11 in [19, Section 7.3]. The manifold optimization toolbox Manopt [34] implements a variation of this algorithm. We use this open-source toolbox Manopt in our implementation of Algorithm 1 by leveraging its flexibility for selectable choices of stopping criteria, tolerances, and other parameters.

IV. Convergence and Analysis

A. Convergence Conditions and Properties of CMA

The global convergence properties of CMA for a single PAM or QAM source recovery in noiseless scenarios are well known [3, Chapters 4-7]. The case of SIMO-CMA blind equalizers, also known as fractionally-spaced CMA or CMA-FSE (when applied to blind equalization scenarios), correspond to the case of recovering one transmitted signal in a grant-free scenario. The CMA-FSE equalizer has guaranteed global convergence to an optimum so long as the channel matrix $\boldsymbol{H}$ has full column rank [35]. Additionally, in [36] the authors show that in the asymptotic regime and under the aforementioned assumptions, for all stable critical points (desirable or not) the only direction of invariance of the CMA cost function corresponds to the 1-dimensional manifold of rotations of the combiner, or $\mathcal{U}(1)$ using our notation in this paper. This invariance result is exploited in [37] to further establish conditions for convergence of CMA and adapt a Newton method, discarding search directions in the equivalence class of rotations.

Multi-channel CMA equalizers are an extension of CMA-FSE, where multiple transmitter simultaneously send independent source signals on a set of shared channels. Targeting one source signal, we can apply CMA to update a receiver equalizer that recovers the signal with minimal co-channel multi-user interference and minimum inter-symbol interference (ISI) [17], [38]. Global convergence of MIMO-CMA equalizers have similar requirements as the case of CMA-FSE equalizers, which in turn is equivalent to have the channel convolution matrix $\boldsymbol{H}$ with full-column rank. Thus, channel matrix $\boldsymbol{H}$ of full-column rank provides guaranteed global convergence to an optimum solution (minimum) in noiseless scenarios. Multiple source recovery as presented here is a special case of the multiple source recovery scheme presented in [17] with zero-ISI subchannels. Thus, global convergence to an optimum solution is also guaranteed under similar conditions. The effect of moderate channel noises on CMA has been shown to be mild by introducing additional local minima in the vicinity of the global solution [4], [39]. Hence, in the following we will always assume that the channel matrix $\boldsymbol{H}$ has full-column rank.

However, most of these works focus in the asymptotic behavior of the CMA cost function and its geometry. Under finite data samples, our recent work [40] provides the first analysis for CMA-based source recovery. Our findings in [40] show that despite being non-convex, the CMA cost function is geometrically well-behaved in terms of strong convexity and bounded curvature in a neighborhood of the optimal solutions. The CMA convergence is therefore tractable in the non-asymptotic case. In the remainder of this section, we shall utilize some of the recent results to bound the minimum eigenvalue of the Euclidean Hessian, in order to provide convergence guarantees for RTR.

B. Known Results in Relation to ROCMA

Riemannian manifold optimization with different solvers, such as RTR, has well-known convergence guarantees [33] over several classical manifolds, such as the Stiefel manifold and the generalized Stiefel manifold, [29], the Grassmannian manifold [19], and many others. These properties also apply to quotient Riemannian manifolds, as the computational space is still the ambient manifold [19]. In particular, RTR will converge superlinearly to nondegenerate local minima and avoid other stationary points in practice, owing to its descent nature [33]. Moreover, the set of stationary points of the Riemannian gradient of CMA is the same as the set of stationary points in Euclidean space, and no other critical points are introduced under the manifold constraints.

Previous works have analyzed some particular cases of cost functions that are mathematically similar to Eq.(18). These works present scenarios closely related to the constant modulus portion of the OCMA problem, but did not exploit the (scaled) orthogonality of several solutions in the problem geometry. In [41] the authors optimize over the Stiefel manifold to maximize the diagonal terms of a matrix quadratic form for joint diagonalization, which is similar to the CMA cost function by setting $R_2 = 0$. Another work [42] tackles the phase retrieval problem by defining a manifold geometry with the so-called fixed-norms manifold.

We now adopt and extend existing analysis to investigate the convergence properties of ROCMA.

C. Convergence of ROCMA

We first make slight modifications to simplify analysis. Recall the SVD of the full rank channel matrix H from Section II-D. We can rewrite

$$\boldsymbol{W}^{\mathsf{H}}\boldsymbol{x}_k = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{z}_k = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{V}^{\mathsf{H}}\boldsymbol{s}_k + \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{n}_k = \boldsymbol{Q}^{\mathsf{H}}\boldsymbol{s}_k + \breve{\boldsymbol{n}}_k,$$

where we use $\breve{\boldsymbol{n}}_k = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{n}_k$ to denote the demixer output noise vector. Note that here we use $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{Y} = \boldsymbol{H}^{\mathsf{H}}\boldsymbol{W} \in \mathbb{C}^{L\times J}$ to denote the matrix of combined (channel plus demixer) parameter vectors. This linear variable transformation between $\boldsymbol{Q}$ and $\boldsymbol{Y}$ is bijective since $\boldsymbol{H}$ is full-column rank and $\boldsymbol{V}$ is orthonormal. Hence, our analysis will equivalently be performed in $\boldsymbol{Q}$-space and $\boldsymbol{Y}$-space. Without loss of generality, we consider noiseless scenario for ease of exposition.

As $\boldsymbol{Q}$ interacts with independent source signals directly, it is straightforward to check whether CMA converges to a joint demixer of the form $\widehat{\boldsymbol{Q}} = \boldsymbol{P}\boldsymbol{D}$, in which $\boldsymbol{P} \in \mathbb{C}^{L\times J}$ is a tall permutation matrix, and $\boldsymbol{D}$ a diagonal unitary matrix. Hence, any optimum solution of ROCMA in the quotient manifold $\mathcal{M} = \mathrm{ST}(L,J)/\mathcal{U}(1)^{\times J}$ has to be of the form $[\widehat{\boldsymbol{Y}}] = [\boldsymbol{V}^{\mathsf{H}}\boldsymbol{P}]$. Equivalently, in the ambient manifold, the horizontal lift of the optimum is $\widehat{\boldsymbol{Y}} = \mathrm{lift}_{[\widehat{\boldsymbol{Y}}]}([\widehat{\boldsymbol{Y}}])$. In the following, we adopt all these representations equivalently.

Recall that the second order moment, fourth order moment, and kurtosis of QAM source signals are defined [3] by

$$m_2 = \mathbb{E}\{|s[k]|^2\},\ m_4 = \mathbb{E}\{|s[k]|^4\},\ \kappa = m_4 - 2m_2^2 < 0.$$

We note here that for QAM signals, $4m_2^2+3\kappa \geq m_2^2$ holds. We now present the two main theorems that ensure convergence guarantees for ROCMA using RTR.

Theorem 1. *Consider signal vector $\boldsymbol{s}_k \in \mathbb{C}^L$ with i.i.d. elements from a square QAM constellation of size Q. Additionally, let the channel matrix $\boldsymbol{H}$ be full-column rank. Let $[\widehat{\boldsymbol{Y}}]$ be a solution of the ROCMA problem on the quotient manifold $\mathcal{M} = \mathrm{ST}(L,J)/\mathcal{U}(1)^{\times J}$. There exist $C_1 > 0$ and $c_1 > 0$ such that, if the number of measurements $K \geq C_1 L$, the Riemannian Hessian of the cost function $g(\cdot)$ in $\mathcal{M}$ is positive definite in the neighborhood of $[\widehat{\boldsymbol{Y}}]$ with probability of at least $1 - 6\mathrm{e}^{-c_1 K}$.*

Proof. We begin our proof from the case of single-source recovery ($J = 1$) before generalizing to multiple sources $J > 1$.

As the cost function is a sample average, we study the expectation of its Hessian and invoke concentration of measure to approximate the behavior of the Hessian with that of its mean with high probability. Let $\widehat{\boldsymbol{Q}} \in \mathbb{C}^L$ be an ROCMA minimum in combined parameter space. Specifically, $\widehat{\boldsymbol{Q}} = \mathrm{e}^{j\theta}\boldsymbol{e}_i$ which denotes arbitrary θ rotation of a canonical vector $\boldsymbol{e}_i$. Invoking [40, Lemma 1], for $\delta > 0$ there exist $C_1(\delta) > 0$ and $c_1(\delta) > 0$ such that for $K \geq C_1(\delta)L$, the single-source sample Euclidean Hessian is concentrated around its mean, i.e., $\left\|\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}}) - \mathbb{E}\{\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}})\}\right\| \leq \delta$ with probability at least $1 - 6\mathrm{e}^{-c_1(\delta)K}$.

Let $\boldsymbol{U} \in \mathbb{C}^L$ a nonzero vector. From [40], the quadratic form of the mean single-source Euclidean Hessian at $\widehat{\boldsymbol{Q}}$ (using Wirtinger derivatives as in [37]) is

$$\begin{aligned}
&\frac{1}{2}\begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}^{\mathsf{H}} \mathbb{E}\{\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}})\}\begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}\\
&\quad= (2m_2^2 - R_2 m_2)\|\boldsymbol{U}\|^2 + (3m_2^2 + 2\kappa)\mathrm{Re}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2\\
&\qquad - m_2^2\mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2 + (m_2^2 + \kappa)|\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}}|^2\\
&\quad= |\kappa|\cdot\|\boldsymbol{U}\|^2 + (4m_2^2 + 3\kappa)\mathrm{Re}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2 + \kappa\mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2.
\end{aligned} \tag{39}$$

Additionally, knowing that $4m_2^2 + 3\kappa \geq m_2^2$ and

$$\mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2 \leq \max|U_i|^2 \leq \sum_{i=1}^{L}|U_i|^2 = \|\boldsymbol{U}\|^2, \tag{40}$$

we see that $\boldsymbol{U}$ is a zero solution of Eq.(39) if and only if $\boldsymbol{U}$ has one nonzero element corresponding to the nonzero element of $\widehat{\boldsymbol{Q}}$, and $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{Q}$ is purely imaginary, i.e. $\|\boldsymbol{U}\|^2 = \mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2$. Let $\widehat{\boldsymbol{Y}} \in \overline{\mathcal{M}} = \mathrm{ST}(L,1)$ be the horizontal lift of the optimum $[\widehat{\boldsymbol{Y}}]$. Recall that $\widehat{\boldsymbol{Y}}^H\widehat{\boldsymbol{Y}} = 1$. From the relationship between $\overline{\mathcal{M}}$ and combined parameter space, i.e. $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{Y}$ with the matrix $\boldsymbol{V}$ formed by right singular vectors of $\boldsymbol{H}$, we have that the optimum in combined parameter space is $\widehat{\boldsymbol{Q}} = \boldsymbol{V}\widehat{\boldsymbol{Y}}$. Now let $\boldsymbol{G} = \boldsymbol{V}^{\mathsf{H}}\boldsymbol{U}$, which we decompose in three orthogonal components: a vertical direction $\boldsymbol{E} = \widehat{\boldsymbol{Y}}\boldsymbol{T} \in \mathrm{V}_{\widehat{\boldsymbol{Y}}}$ with $\boldsymbol{T}$ imaginary diagonal (i.e., a scalar), a horizontal direction $\boldsymbol{F} \in \mathrm{H}_{\widehat{\boldsymbol{Y}}}$ and a normal direction $\boldsymbol{C} = \widehat{\boldsymbol{Y}}\boldsymbol{A} \in \mathrm{N}_{\widehat{\boldsymbol{Y}}}\overline{\mathcal{M}}$, where $\boldsymbol{A}$ is Hermitian (i.e., a real scalar). Hence,

$$\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}} = \boldsymbol{G}^{\mathsf{H}}\widehat{\boldsymbol{Y}} = \boldsymbol{E}^{\mathsf{H}}\widehat{\boldsymbol{Y}} + \boldsymbol{F}^{\mathsf{H}}\widehat{\boldsymbol{Y}} + \boldsymbol{C}^{\mathsf{H}}\widehat{\boldsymbol{Y}} = \boldsymbol{T}^{\mathsf{H}} + \boldsymbol{A}, \tag{41}$$

where we use that $\boldsymbol{F}^{\mathsf{H}}\widehat{\boldsymbol{Y}} = 0$ because it is skew-Hermitian with zero diagonal and a scalar. Now, given $\boldsymbol{U}$ for which $\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}}$ is purely imaginary, the Hermitian $\boldsymbol{A}$ must be 0. In conclusion,

the *only* directions in the nullspace of the Euclidean Hessian at $\widehat{\boldsymbol{Q}}$ in combined parameter space correspond only to vertical directions at $\widehat{\boldsymbol{Y}}$ in $\boldsymbol{Y}$-space, i.e., $\mathrm{V}_{\widehat{\boldsymbol{Y}}}$ corresponds to the nullspace of the Hessian. Therefore, in the quotient manifold $\mathcal{M}$, where the vertical directions are eliminated, the mean Hessian must be positive definite on the manifold.

Now, according to [40, Lemma 1], $K > C_1(\delta)L$ implies that $\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}}) \succeq \mathbb{E}\{\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}})\} - \delta\boldsymbol{I}$ with probability of at least $1 - 6\mathrm{e}^{-c_1 K}$. From Eq.(39) we have

$$\begin{aligned}
&\frac{1}{2}\begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}^{\mathsf{H}} \nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}}) \begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}\\
&\geq \frac{1}{2}\begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}^{\mathsf{H}} \mathbb{E}\big\{\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}})\big\} \begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix} - \frac{\delta}{2}\begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}^{\mathsf{H}} \begin{bmatrix}\boldsymbol{U}\\ \overline{\boldsymbol{U}}\end{bmatrix}\\
&\geq (|\kappa| - \delta)\|\boldsymbol{U}\|^2 + (4m_2^2 + 3\kappa)\mathrm{Re}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2 + \kappa\mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2,
\end{aligned}$$

which is strictly positive for $\delta < |\kappa|\big(\|\boldsymbol{U}\|^2 - \mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2\big)$, where $\mathrm{Im}(\boldsymbol{U}^{\mathsf{H}}\widehat{\boldsymbol{Q}})^2 < \|\boldsymbol{U}\|^2$ because $\boldsymbol{G} = \boldsymbol{V}^{\mathsf{H}}\boldsymbol{U}$ does not belong to the vertical space $\mathrm{V}_{\widehat{\boldsymbol{Y}}}$.

We now generalize to the multiple source recovery case ($J > 1$). Let $\widehat{\boldsymbol{Q}}$ be an optimum, that is, $\widehat{\boldsymbol{Q}} = \boldsymbol{P}\boldsymbol{D}$, with $\widehat{\boldsymbol{Q}}_\ell = \mathrm{e}^{j\theta_\ell}\boldsymbol{e}_{i_\ell}$ its ℓ-th column corresponding to the ℓ-th optimum demixer. The Euclidean cost function $\overline{\overline{g}}$, written in terms of each demixer similar to Eq.(6) contains no cross-product between demixing combiners. Through vectorization of $\widehat{\boldsymbol{Q}}$ and permutations of columns and rows, we obtain a Wirtinger Hessian matrix such that its elements are ordered for each combiner as in [37], [40]. Thus, the Hessian matrix is block diagonal, with each block being the single-source Euclidean Hessian evaluated at the ℓ-th combiner, i.e. $\nabla^2\overline{\overline{g}}(\boldsymbol{Q}_\ell)$, $\ell \in \{1, \cdots, J\}$.

Let $\boldsymbol{U} \in \mathbb{C}^{L\times J}$ a nonzero matrix, and $\boldsymbol{U}_\ell$ its ℓ-th column. In quadratic form, the mean Euclidean Hessian at $\widehat{\boldsymbol{Q}}$ can be rewritten as

$$\sum_{\ell=1}^{J} \frac{1}{2}\begin{bmatrix}\boldsymbol{U}_\ell\\ \overline{\boldsymbol{U}_\ell}\end{bmatrix}^{\mathsf{H}} \mathbb{E}\big\{\nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{Q}}_\ell)\big\} \begin{bmatrix}\boldsymbol{U}_\ell\\ \overline{\boldsymbol{U}_\ell}\end{bmatrix}, \tag{42}$$

where each summand is derived from Eq.(39). Following the same reasoning as in the single-source case, $\boldsymbol{U}_\ell$ is a zero solution of the ℓ-th summand of Eq.(42) if and only if $\|\boldsymbol{U}_\ell\|^2 = \mathrm{Im}(\boldsymbol{U}_\ell^{\mathsf{H}}\widehat{\boldsymbol{Q}}_\ell)^2$. Furthermore, $\boldsymbol{U}_\ell$ has to correspond to the ℓ-th column of a vertical direction at $\widehat{\boldsymbol{Y}}$ in $\boldsymbol{Y}$-space.

Hence, $\boldsymbol{U} = \boldsymbol{V}\widehat{\boldsymbol{Y}}\boldsymbol{T}$ with $\boldsymbol{T}$ imaginary diagonal. Noting that $\widehat{\boldsymbol{Y}}$ has orthogonal columns by definition, and that $\boldsymbol{T}$ has J free parameters, the nullspace of the Hessian has dimension J. Again we conclude that the only directions in the nullspace of the Hessian at $\widehat{\boldsymbol{Q}}$ correspond to vertical directions at $\widehat{\boldsymbol{Y}}$. As we ignore vertical directions when g is restricted to the quotient manifold $\mathcal{M}$, the mean Hessian must be positive definite.

Now, due to the block diagonal structure of the Hessian, it is straightforward to bound the quadratic form of the sample Hessian as follows:

$$\begin{aligned}
\sum_{\ell=1}^{J} \frac{1}{2}\begin{bmatrix}\boldsymbol{u}\\ \overline{\boldsymbol{u}}\end{bmatrix}^{\mathsf{H}} \nabla^2\overline{\overline{g}}(\widehat{\boldsymbol{q}}_\ell) \begin{bmatrix}\boldsymbol{u}\\ \overline{\boldsymbol{u}}\end{bmatrix} \geq &\sum_{\ell=1}^{J}(|\kappa| - \delta)\|\boldsymbol{u}_\ell\|^2 + \kappa\mathrm{Im}(\boldsymbol{u}_\ell^{\mathsf{H}}\widehat{\boldsymbol{q}}_\ell)^2\\
&+ (4m_2^2 + 3\kappa)\mathrm{Re}(\boldsymbol{u}_\ell^{\mathsf{H}}\widehat{\boldsymbol{q}}_\ell)^2,
\end{aligned}$$

which again is strictly positive for δ small enough with high probability, following the same procedure as above. ■

Theorem 2 (Convergence of RTR in ROCMA)**.** *Under the conditions of Theorem 1, ROCMA is globally and locally convergent in $\mathcal{M}$ with high probability using Riemannian Trust Regions as defined in Algorithm 1.*

Proof. First, we prove global convergence. The Riemannian quotient manifold $\mathcal{M}$ is smooth and it is compact since it is a quotient space of $\mathrm{ST}(L, J)$, which itself is compact. The cost function g is smooth because its lifted version is a (smooth) matrix polynomial. Invoking [33, Corollary 4.6], ROCMA satisfies all conditions for global convergence of RTR.

For local convergence, we prove that [33, Theorem 4.12] holds for ROCMA. In particular, ROCMA needs to meet the following requirements:

- All conditions of [33, Theorem 4.2], which are satisfied owing to global convergence.
- The retraction needs to satisfy the bound [33, Eq.(18)], which holds because: (a) $\mathcal{M}$ is compact, and (b) the lifted retraction is smooth.
- The norm of the inverse Hessian operator needs to be bounded in a neighborhood of the optimum. Invoking Theorem 1, with high probability the Riemannian Hessian in $\mathcal{M}$ is strictly positive definite with minimum eigenvalue $\lambda_{\min} > 0$. Since the Hessian is continuous, its inverse is bounded by $\lambda_{\min}^{-1}$ in a neighborhood of any optimum.

Hence, with high probability, ROCMA enjoys the global and local convergence guarantees of RTR as stated in [33]. ■

D. Computational Complexity

We can estimate the computational complexity of the ROCMA algorithm by analyzing each step of Algorithm 1. In particular, the Riemannian Trust-Region step includes iterations needed for convergence and also iterations of each call to the trust-region subproblem algorithm. Henceforth, we refer to the former as outer iterations, and the latter as inner iterations.

1) The source estimation step is dominated by the computation of Minka's Laplace method, with complexity of $\mathcal{O}(ML)$, and a number of eigendecompositions, to be obtained iteratively with a cost of $\mathcal{O}(M^3)$.
2) The initialization is dominated by obtaining $\boldsymbol{z}_k$ via scaling, with a complexity level of $\mathcal{O}(MLK)$. Other computations have relatively insignificant cost, mainly for defining geometry operations.
3) Each inner iteration is dominated by the computation of the Riemannian Hessian, at complexity $\mathcal{O}(L^2K)$. Other operations are linear over the samples with a cost of $\mathcal{O}(LK)$.
4) Each outer iteration is dominated by the total inner iterations required in the particular outer iteration I_t. Thus, this amounts to complexity of $\mathcal{O}(L^2KI_t)$. The Riemannian retraction $\mathrm{R}_{\boldsymbol{Y}}$ requires $\mathcal{O}(L^3)$, whereas other operations are linear over scalars and of $\mathcal{O}(1)$. The final scaling by $\boldsymbol{U}_L\boldsymbol{\Sigma}_L^{-1}$ has a cost of $\mathcal{O}(ML^2)$.

Based on the aforementioned steps, Table II summarizes the complexity of ROCMA in each step.

TABLE II: Computational complexity of ROCMA.

ROCMA Steps	Total cost
Estimate L	$\mathcal{O}(ML)$
Iterative eigendecomposition $\boldsymbol{R_Y}$	$\mathcal{O}(M^3)$
Define $\boldsymbol{z}_k$	$\mathcal{O}(MLK)$
RTR outer iteration	$\mathcal{O}(L^3 + L^2KI_t)$
Final scaling	$\mathcal{O}(ML^2)$

In comparison, the computational complexity of a typical gradient-descent scheme in Euclidean space is of $\mathcal{O}(M^2LK)$ per iteration [40], which is lower than the cost of RTR iterations for a moderate number of inner iterations. The iteration complexity of RTR can be found in [31, Theorem 6.10]. In our numerical results of Section V, we shall consider some benchmark applications and provide a computation comparison of RTR with other Riemannian and traditional CMA algorithms.

V. Simulation Results

A. Definitions

We consider a multi-user signal recovery scenario consisting of L source nodes, each transmitting K independent symbols from a typical QAM (QPSK or 16-QAM) constellation with normalized average power. The serving node has M receive antennas whereas each low cost source node has a single antenna. We model wireless channels $\boldsymbol{H}$ as stationary Rayleigh with i.i.d. entries $H_{ml} \sim \mathcal{N}(0, \frac{1}{2}) + i\mathcal{N}(0, \frac{1}{2})$, in addition to additive white Gaussian noise vector $\boldsymbol{n}_k$ with i.i.d. entries and variance (power) σ^2 consistent with receiver SNR of 20dB. To measure signal recovery performance, we evaluate normalized total interference (NTI) for each recovered source, defined by

$$\mathrm{NTI}_\ell = \frac{\sum_i |C_{\ell i}|^2 - \max_i |C_{\ell i}|^2}{\max_i |C_{\ell i}|^2} \tag{43}$$

where $C_{\ell i}$ are the entries of the final mixing matrix $\boldsymbol{C} = \boldsymbol{W}^{\mathsf{H}}\boldsymbol{H}$ after demixing, with its rows corresponding to each equalized channel-demixer pair. Unless otherwise stated, we average 100 runs per setting. For comparison purposes, we implement several recovery algorithms:

- **Riemannian Trust-Regions (RTR):** We use Manopt [34] to implement the geometry of ROCMA, and use the built-in trust regions solver with default initial trust radius of $\varepsilon = 0.125\sqrt{\dim \mathcal{M}}$ and random initialization.
- **Riemannian Gradient Descent (RGD):** We use our ROCMA geometry implementation in Manopt and use the built-in steepest descent solver, which applies batch gradient descent, i.e. sample average gradients. We use the default adaptive linesearch of Manopt and random initialization.
- **Wirtinger Flow CMA (WF):** Another algorithm [40] uses spectral initialization and batch gradient descent updates normalized by the iterate norm. For computational fairness, we implement WF in Manopt (with Euclidean geometry) and use its default adaptive linesearch.
- **Gram-Schmidt CMA (GS)**: We also test the simpler stochastic gradient descent method of [18], which applies one-sample gradient updates and Gram-Schmidt orthogonalization in each iteration to achieve signal separation, for a maximum of 20000 iterations and with small step-size $\mu = 10^{-4}$, which achieves convergence in our tests. The available K signal samples are cyclically reused until termination.

All algorithms aim to recover the same number of sources estimated by Minka's Laplace method [28], and share the same stopping criteria: reaching a maximum number of iterations, or having a gradient with norm smaller than 10^{-12}.

B. Signal Recovery Efficacy

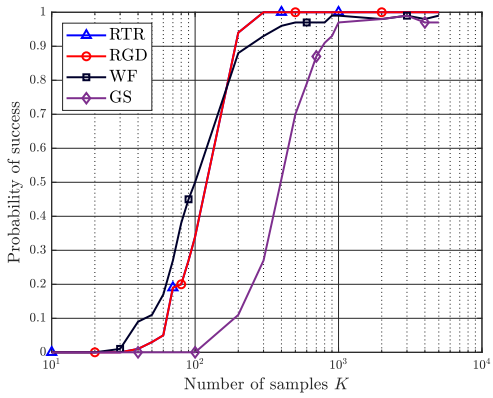
We test and compare the power of source recovery capabilities under different system sizes, different number of samples, and different QAM constellations.

Fig. 3 presents the probability of successful recovery of multiple sources based on different numbers of received data sample, under 20dB of SNR. We test both QPSK and 16-QAM modulations for system sizes of $M \times L = 8 \times 4$ and 16×8. We define *success* as the event that *all* source signals are recovered with NTI below -20dB.

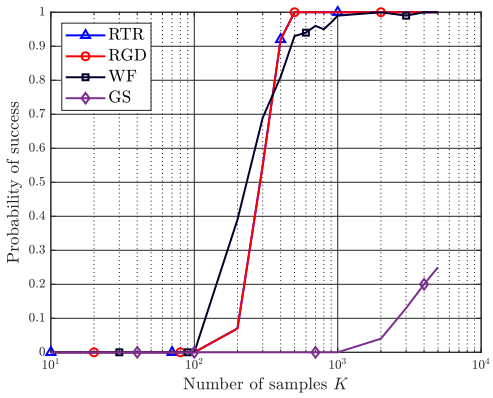
Fig. 3a shows a 8×4 system using QPSK source modulation. Our proposed RTR and RGD exhibit nearly the same probability of success for various sample size K, both successfully achieving 100% success for $K \geq 300$ samples. In comparison, the existing WF algorithm exhibits similar or slightly higher recovery success rate at first, but stagnates and fails to achieve 100% of success with increasing data size. Unlike RTR and RGD, the WF receiver requires at least 900 data samples to achieve a probability of success above 98%. Applying sequential Gram-Schmidt orthogonalization, GS requires even more samples to achieve higher probability of success. In fact, our test fails to reach above 96% success probability despite utilizing thousands of data samples.

Testing a larger system using QPSK, results from Fig. 3b demonstrate similar relative performance by RTR, RGD and WF, requiring 500 and 1000 samples to achieve over 99% probability of success. The GS demixer, however, is much more sensitive to problem size, and struggles to jointly recover all signals even with 5000 received data samples.

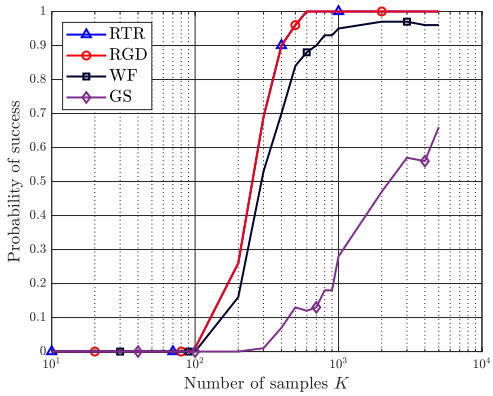
We now consider the more challenging case of 16-QAM source signals. Figs. 3c and 3d respectively provide the results for problem size of 8×4 and 16×8. The test results show that the two Riemannian methods (RTR and RGD) again achieve high probability of successful signal recovery, now requiring only 600 and 2000 samples respectively for near 100% success rate. The WF algorithm is less successful in comparison, requiring at least 2000 samples to achieve more than 95% of success probability. On other other hand, GS only achieves 60% probability of success for 16-QAM for the smaller 8×4 system size. For the more complex system of size 16×8, GS is in fact unable to recover all 8 detected sources under the stated simulation settings. Indeed, this phenomenon was anticipated by [18]. Since GS relies on sequential Gran-Schmidt orthogonalization for signal separation, error propagation can lead to weaker interference suppression. As the ℓ-th demixer relies on all previous demixers in order to isolate a different signal,



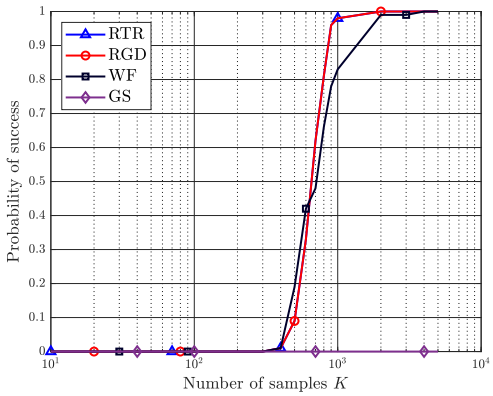
(a) QPSK, $M \times L = 8 \times 4$.



(b) QPSK, $M \times L = 16 \times 8$.



(c) 16-QAM, $M \times L = 8 \times 4$.



(d) 16-QAM, $M \times L = 16 \times 8$.

Fig. 3: Probability of successful recovery of all detected demixers vs. number of samples K.

GS becomes increasingly more vulnerable to error propagation with increasing number of signals.

Overall, both proposed Riemannian algorithms (RTR and RGD) demonstrate strong performance across different modulations and system sizes under test than their Euclidean counterparts. WF still shows high probability of successful recovery, albeit at the cost of more samples than RTR or RGD. GS, on the other hand, struggles against moderately large system sizes or higher-order modulations.

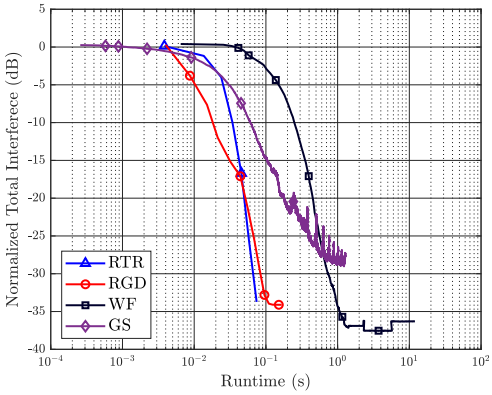
C. Computation and Interference Rejection

For the purposes of signal recovery, computation complexity must be jointly analyzed with respect to the achieved level of interference rejection in recovered signals. For this reason, in the following presentation we show how well our proposed methods (RTR and RGD) work in terms of both interference rejection and computational load. We also provide benchmark comparison with the two algorithms (WF and GS).

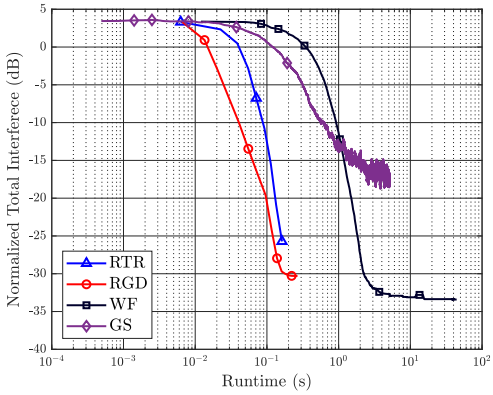
We first investigate algorithm runtime ("wall-clock" time). Runtime depends on computer, platform and code implementation. Thus, runtime comparison alone may not fully capture true algorithm runtime. To mitigate this possible bias, we let all source codes for RTR, RGD and WF utilize the same Manopt structure and libraries. GS, on the other hand, uses a direct implementation using the fastest routines available (matrix products, etc.) for each iteration.

Fig. 4 depicts the average interference rejection over all runs of the algorithms with respect to their runtime, for a fixed number of samples $K = 2000$. From the results of all tested modulation schemes and system sizes, both RGD and RTR, by exploiting the proposed geometry, are faster than GS by at least 4 times, and are faster than WF by an order of magnitude. Such advantage is consistent across nearly all tested modulation schemes and system sizes. Moreover, RTR and RGD exhibit similar runtime across modulations, and require about double the time to converge in the larger system of 16×8, which contains 2 times more sources and antennas (Figs. 4b and 4d). There are some modest differences between the two Riemannian solvers. In particular, RGD converges slightly faster initially to achieve moderate interference mitigation of approximately -15dB. For higher interference rejection, both RGD and RTR show similar runtime complexity.

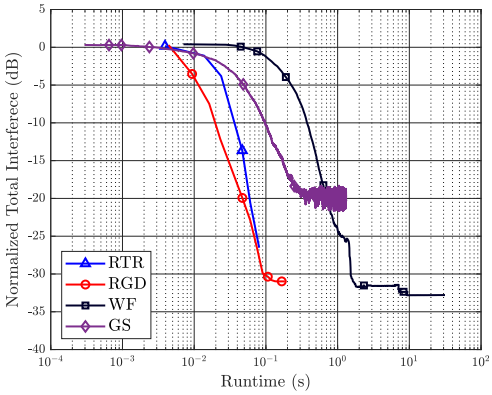
WF exhibits a similar efficacy in terms of interference rejection, but it does require longer runtime to achieve a particular level of interference rejection. On the other hand, GS requires more runtime to converge than the Riemannian methods across all tested scenarios. Moreover, GS is less effective in interference mitigation, particularly for more complex source signals of 16-QAM, as shown in Fig. 4c and Fig. 4d. From the test results, it is evident that GS stalls in terms of interference rejection and additional iterations do not improve the performance of interference rejection. This phenomenon is



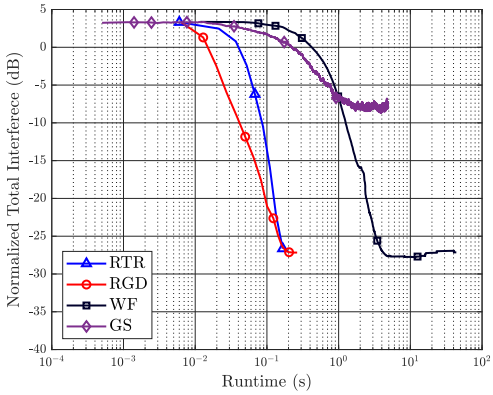
(a) QPSK, $M \times L = 8 \times 4$.



(b) QPSK, $M \times L = 16 \times 8$.



(c) 16-QAM, $M \times L = 8 \times 4$.



(d) 16-QAM, $M \times L = 16 \times 8$.

Fig. 4: Average total interference for all detected demixers vs. runtime.

consistent with the discussion of error propagation in sequential orthogonalization of the GS approach. The performance loss of GS becomes more severe with increasing modulation and system size, e.g. is about -30dB of interference for QPSK signals and a 8×4 system (Fig. 4a), but is less than -10dB for 16QAM and 16×8 system (Fig. 4d).

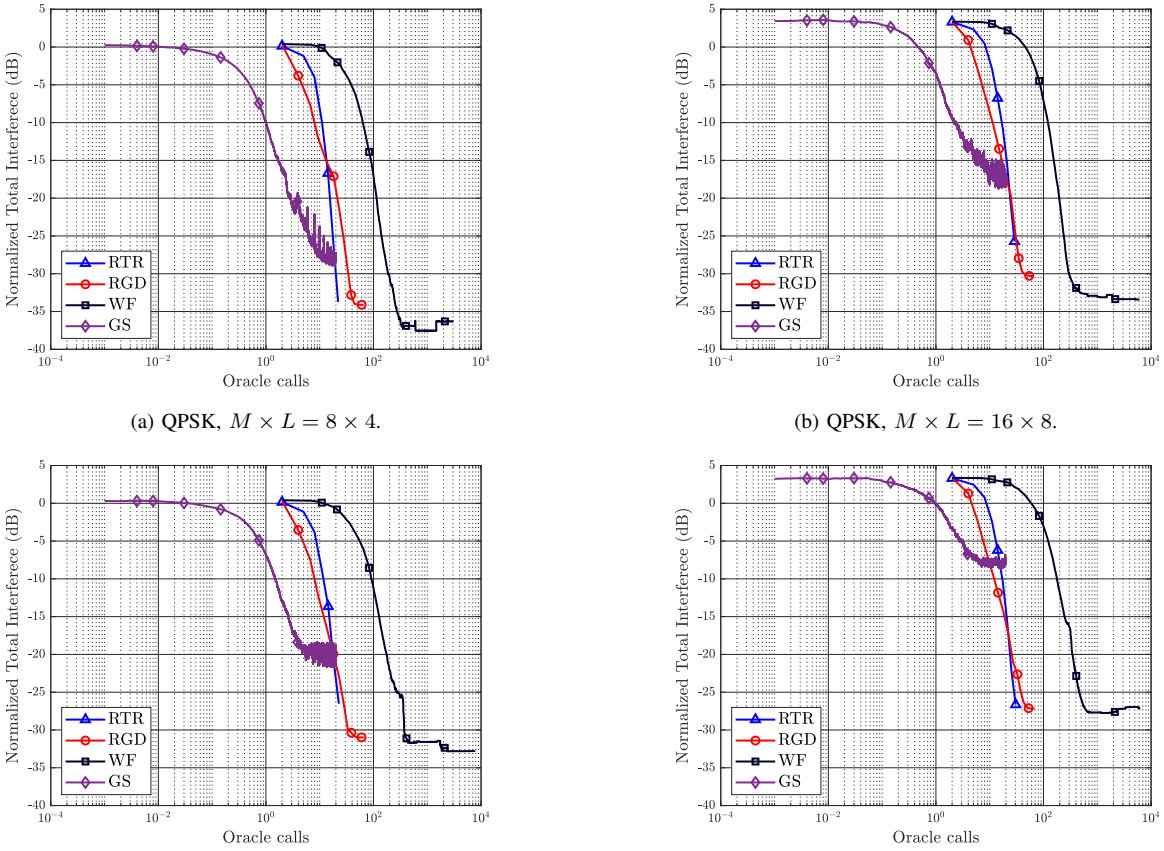
To address the potential bias of over-reliance on runtime when analyzing computation complexity, we also compare the algorithm behavior respect to *oracle calls*, that is, the total number of cost function, gradient, Hessian and GS orthogonalization calls/evaluations of each algorithm. These operations are dominant contributors to total computational complexity, and hence they act as a proxy for platform-independent computational load assessment. Recall that GS uses only one sample per gradient update and its orthogonalization does not increase with the number of samples, we normalize the number of oracles calls by K in order to provide a fair comparison with the other methods, that use sample averages in all its oracle calls.

Fig. 5 shows the average achieved NTI of each algorithm with respect to their oracle calls, for a fixed number of samples $K = 2000$. As seen in previous comparisons, RTR and RGD have similar computational cost. Across the tested modulations and system sizes, both RGD and RTR require at least an order of magnitude fewer function computations than WF to achieve 20dB of interference mitigation. Even if we account for their larger computational complexity per iteration than WF or GS, the total cost of every algorithm is dominated by the number of iterations, and RTR and RGD incur significantly lower total computational load than WF. WF shows good interference rejection power, but requires about 10-20 times more oracle computations than the Riemannian methods to achieve equal levels of interference mitigation.

As expected, GS is the least complex algorithm owing to its simple stochastic gradient update using only one sample in each iteration. At an average interference suppression of 20dB, the computational cost of GS is about 5-8 times lower than the Riemannian methods, as seen in Figs. 5a-5c. However, the sequential nature of GS leads to poorer interference rejection, particularly for higher-dimensional QAM constellation and/or larger system sizes. In comparison, the proposed Riemannian algorithms can achieve substantially better performance of signal recovery with only a modest increase of computation complexity. Exploiting Riemannian geometry provides a good complexity-efficacy tradeoff.

VI. Conclusion

In this paper, we present a new approach for blind multiple signal recovery based on Riemannian geometry. Formulating Constant-Modulus Algorithm into a novel Riemannian optimization problem, we develop a direct algorithm to achieve



(a) QPSK, $M \times L = 8 \times 4$.

(b) QPSK, $M \times L = 16 \times 8$.

(c) 16-QAM, $M \times L = 8 \times 4$.

(d) 16-QAM, $M \times L = 16 \times 8$.

Fig. 5: Average total interference for all detected demixers vs. oracle calls.

blind recovery of multiple source signals by embedding independent signal demixers within the geometry of a Riemannian manifold. In our work, we derive the novel geometry and the formulation that make it possible to minimize the constant modulus cost function over the Riemannian manifold as the search space of the optimization problem. In particular, the proposed quotient manifold geometry allows for a positive definite Hessian operator at optimum solutions and provides stronger theoretical convergence guarantees while bypassing the cost function invariance problem. Our approach demonstrates high probability of successful recovery of all detected sources with a moderate number of samples, for practical system sizes and different modulation schemes. Furthermore, the proposed Riemannian solutions present a good tradeoff between computation complexity and interference suppression.

Future research paths include multiple source recovery and equalization over ISI fading channels with memory, stochastic or mini-batch reformulation for complexity reduction, and the definition of new geometrical perspectives that could exploit other known signal information such as forward error correction code within the optimization framework.

REFERENCES

[1] D. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, November 1980.

[2] J. Treichler and M. Larimore, "New Processing Techniques Based on the Constant Modulus Adaptive Algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 420–431, 1985.

[3] Z. Ding and Y. G. Li, *Blind Equalization and Identification*. New York, NY, USA: Marcel Dekker, Inc., December 2000.

[4] H. H. Zeng, L. Tong, and C. R. Johnson, "An Analysis of Constant Modulus Receivers," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 2990–2999, Nov 1999.

[5] O. Dabeer and E. Masry, "Convergence Analysis of the Constant Modulus Algorithm," *IEEE Transactions on Information Theory*, vol. 49, no. 6, pp. 1447–1464, June 2003.

[6] B. Mariere, Z.-Q. T. Luo, and T. N. Davidson, "Blind Constant Modulus Equalization via Convex Optimization," *IEEE Trans. Signal Processing*, vol. 51, pp. 805–818, 2003.

[7] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[8] K. Wang, "Semidefinite Relaxation Based Blind Equalization using Constant Modulus Criterion," August 2018. [Online]. Available: arXiv:1808.07232v2[cs.IT]

[9] A. Adler and M. Wax, "Constant Modulus Algorithms via Low-rank Approximation," *Signal Processing*, vol. 160, pp. 263 – 270, 2019.

[10] B. Hassibi, A. Paulraj, and T. Kailath, "On a Closed Form Solution to the Constant Modulus Factorization Problem," in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, vol. 2, Oct 1994, pp. 775–779.

[11] A.-J. van der Veen and A. Paulraj, "An Analytical Constant Modulus Algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[12] A.-J. van der Veen, "An Adaptive Version of the Algebraic Constant Modulus Algorithm [Blind source separation applications]," in *Proceedings of ICASSP '05. IEEE International Conference on Acoustics,*

Speech, and Signal Processing, vol. 4, March 2005, pp. iv/873–iv/876 Vol. 4.

[13] V. Zarzoso and P. Comon, "Optimal Step-Size Constant Modulus Algorithm," *IEEE Transactions on Communications*, vol. 56, no. 1, pp. 10–13, January 2008.

[14] J. J. Shynk and R. P. Gooch, "Performance Analysis of the Multistage CMA Adaptive Beamformer," in *Proceedings of MILCOM '94*, Oct 1994, pp. 316–320 vol.2.

[15] D. Liu and L. Tong, "An Analysis of Constant Modulus Algorithm for Array Signal Processing," *Signal Processing*, vol. 73, no. 1, pp. 81 – 104, 1999.

[16] A. Bessios and C. Nikias, "Multichannel Adaptive Blind Equalization with CRIMNO-MSE Technique," in *MILCOM 92 Conference Record*, 1992, pp. 236–240 vol.1.

[17] Ye Li and K. J. R. Liu, "Adaptive Blind Source Separation and Equalization for Multiple-Input/Multiple-Output Systems," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2864–2876, Nov 1998.

[18] T. Nguyen and Z. Ding, "CMA Beamforming for Multipath Correlated Sources," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1997, pp. 2521–2524 vol.3.

[19] P.-A. Absil and R. Mahony and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008.

[20] K. Yang, Y. Shi, and Z. Ding, "Low-Rank Matrix Completion for Mobile Edge Caching in Fog-RAN via Riemannian Optimization," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[21] H. Sato, "Riemannian Conjugate Gradient Method for Complex Singular Value Decomposition Problem," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 5849–5854.

[22] W. Huang, K. A. Gallivan, and X. Zhang, "Solving PhaseLift by Low-rank Riemannian Optimization Methods," *Procedia Computer Science*, vol. 80, pp. 1125 – 1134, 2016.

[23] J. Dong, K. Yang, and Y. Shi, "Blind Demixing for Low-Latency Communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 897–911, Feb 2019.

[24] A. Cherian and S. Sra, "Riemannian Dictionary Learning and Sparse Coding for Positive Definite Matrices," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 12, pp. 2859–2871, 2017.

[25] 3GPP TS 36.212 v13.2.0, *Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding*, 3rd Generation Partnership Project, Technical Specification Group Radio Access Network Std., 2016.

[26] A. Jalali and Z. Ding, "Joint Detection and Decoding of Polar Coded 5G Control Channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2066–2078, 2020.

[27] A. Ikhlef and D. Le Guennec, "A Simplified Constant Modulus Algorithm for Blind Recovery of MIMO QAM and PSK Signals: A Criterion with Convergence Analysis," *EURASIP Journal on Wireless Communications and Networking*, vol. 56, no. 6, December 2007.

[28] T. P. Minka, "Automatic Choice of Dimensionality for PCA," M.I.T. Media Laboratory, Cambridge, MA, Tech. Rep. 504, December 2000, (revised Sept. 2008).

[29] A. Edelman, T. A. Arias, and S. T. Smith, "The Geometry of Algorithms with Orthogonality Constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, p. 303–353, Apr. 1999.

[30] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, 1st ed. New York, NY, USA: Springer, 2016.

[31] N. Boumal, "An Introduction to Optimization on Smooth Manifolds," To appear with Cambridge University Press, March 2022. [Online]. Available: http://www.nicolasboumal.net/book

[32] T. A. Palka and R. J. Vaccaro, "Asymptotically Efficient Estimators for Multidimensional Harmonic Retrieval Based on the Geometry of the Stiefel Manifold," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 1691–1695.

[33] P. A. Absil, C. G. Baker, and K. A. Gallivan, "Trust-Region Methods on Riemannian Manifolds," *Found. Comput. Math.*, vol. 7, no. 3, pp. 303–330, Jul. 2007.

[34] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab Toolbox for Optimization on Manifolds," *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014.

[35] Ye Li and Zhi Ding, "Global Convergence of Fractionally Spaced Godard Equalizers," in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, vol. 1, Oct 1994, pp. 617–621 vol.1.

[36] Z. Ding, R. Kennedy, B. Anderson, and C. Johnson, "Ill-convergence of Godard Blind Equalizers in Data Communication Systems," *IEEE Transactions on Communications*, vol. 39, no. 9, pp. 1313–1327, 1991.

[37] K. Kreutz-Delgado and Y. Isukapalli, "Use of the Newton Method for Blind Adaptive Equalization Based on the Constant Modulus Algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3983–3995, Aug 2008.

[38] S. Mayrargue, "A Blind Spatio-Temporal Equalizer for a Radio-Mobile Channel Using the Constant Modulus Algorithm (CMA)," in *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. iii, April 1994, pp. III/317–III/320 vol.3.

[39] H. H. Zeng, L. Tong, and C. R. Johnson, "Relationships Between the Constant Modulus and Wiener Receivers," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1523–1538, 1998.

[40] C. Feres and Z. Ding, "Wirtinger Flow Meets Constant Modulus Algorithm: Revisiting Signal Recovery for Grant-Free Access," *IEEE Transactions on Signal Processing*, vol. 69, pp. 6515–6529, 2021, DOI: 10.1109/TSP.2021.3103038.

[41] F. J. Theis, T. P. Cason, and P. A. Absil, "Soft Dimension Reduction for ICA by Joint Diagonalization on the Stiefel Manifold," in *Independent Component Analysis and Signal Separation*, T. Adali, C. Jutten, J. M. T. Romano, and A. K. Barros, Eds. Springer Berlin Heidelberg, 2009, pp. 354–361.

[42] T. Bendory, Y. C. Eldar, and N. Boumal, "Non-Convex Phase Retrieval From STFT Measurements," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 467–484, 2018.